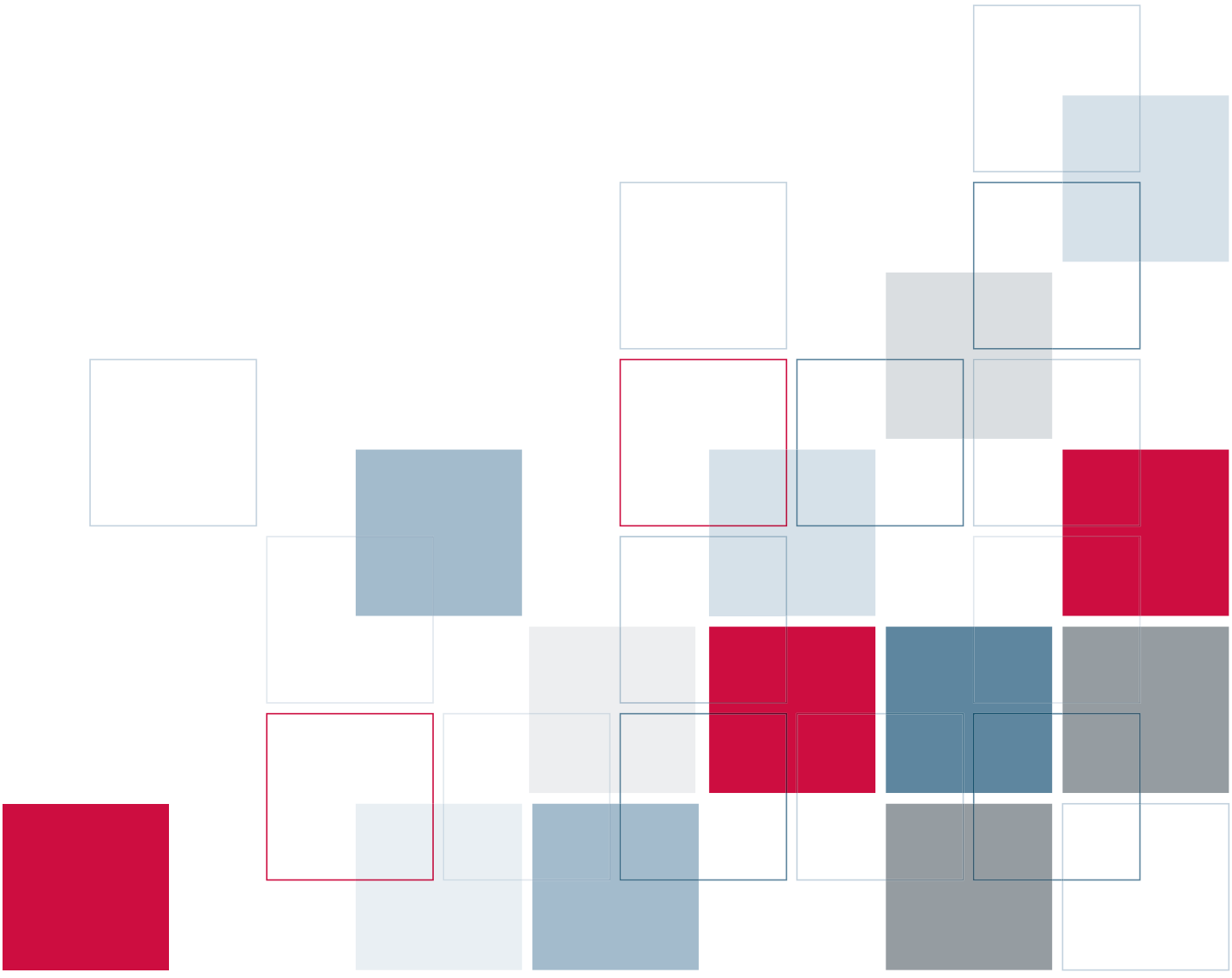
SPSS Base 16.0 – Benutzerhandbuch



Weitere Informationen zu SPSS®-Software-Produkten finden Sie auf unserer Website unter der Adresse *http://www.spss.com* oder wenden Sie sich an

SPSS Inc.
233 South Wacker Drive, 11th Floor
Chicago, IL 60606-6412, USA
Tel.: (312) 651-3000
Fax: (312) 651-3668

SPSS ist eine eingetragene Marke, und weitere Produktnamen sind Marken der SPSS Inc. für Computerprogramme von SPSS Inc. Die Herstellung oder Verbreitung von Materialien, die diese Programme beschreiben, ist ohne die schriftliche Erlaubnis des Eigentümers der Marke und der Lizenzrechte der Software und der Copyrights der veröffentlichten Materialien verboten.

Die SOFTWARE und die Dokumentation werden mit BESCHRÄNKTEN RECHTEN zur Verfügung gestellt. Verwendung, Vervielfältigung und Veröffentlichung durch die Regierung unterliegen den Beschränkungen in Unterabschnitt (c)(1)(ii) von The Rights in Technical Data and Computer Software unter 52.227-7013. Vertragspartner/Hersteller ist SPSS Inc., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.
Patentnr. 7.023.453

Allgemeiner Hinweis: Andere in diesem Dokument verwendete Produktnamen werden nur zu Identifikationszwecken genannt und können Marken der entsprechenden Unternehmen sein.

Windows ist eine eingetragene Marke der Microsoft Corporation.

Apple, Mac und das Mac-Logo sind Marken von Apple Computer, Inc., die in den USA und in anderen Ländern eingetragen sind.

Dieses Produkt verwendet WinWrap Basic, Copyright 1993–2007, Polar Engineering and Consulting, *http://www.winwrap.com*.

SPSS Base 16.0 – Benutzerhandbuch
Copyright © 2007 SPSS Inc.
Alle Rechte vorbehalten.

Ohne schriftliche Erlaubnis der SPSS GmbH Software darf kein Teil dieses Handbuchs für irgendwelche Zwecke oder in irgendeiner Form mit irgendwelchen Mitteln, elektronisch oder mechanisch, mittels Fotokopie, durch Aufzeichnung oder durch andere Informationsspeicherungssysteme reproduziert werden.

ISBN-13: 978-1-56827-832-2
ISBN-10: 1-56827-832-2

1 2 3 4 5 6 7 8 9 0 10 09 08 07

Vorwort

SPSS 16.0

SPSS 16.0 ist ein umfassendes System zum Analysieren von Daten. Mit SPSS können Sie Daten aus nahezu allen Dateitypen entnehmen und aus ihnen Berichte in Tabellenform, Diagramme sowie grafische Darstellungen von Verteilungen und Trends, deskriptive Statistiken und komplexe statistische Analysen erstellen.

Dieses Handbuch, *SPSS Base 16.0 – Benutzerhandbuch*, dokumentiert die grafische Benutzeroberfläche von SPSS. Beispiele für statistische Prozeduren in SPSS Base 16.0 finden Sie im Hilfesystem, das mit der Software installiert wird. Algorithmen für die statistischen Prozeduren sind im PDF-Format über das Menü "Hilfe" verfügbar.

Den Menüs und Dialogfeldern von SPSS liegt eine Befehlssprache zugrunde. Auf einige fortgeschrittene Funktionen des System kann nur mithilfe der Befehlssyntax zugegriffen werden. (Diese Funktionen sind in der Studentenversion nicht verfügbar.) Detaillierte Informationen zur Befehlssyntax sind auf zwei Arten verfügbar: als Bestandteil der umfassenden Hilfesystems und als separates Dokument im PDF-Format im Handbuch *SPSS 16.0 Command Syntax Reference*, das auch über das Menü "Hilfe" verfügbar ist.

SPSS Optionen

Die folgenden Optionen sind als Erweiterungsmodule der Vollversion (nicht der Studentenversion) von SPSS Base erhältlich:

SPSS Regression Models™ bietet Verfahren zur Datenanalyse, die über herkömmliche lineare Statistikmodelle hinausgehen. Es beinhaltet Prozeduren für Probit-Analyse, logistische Regression, Gewichtungsschätzungen, zweistufige Regression kleinster Quadrate und allgemeine nichtlineare Regression.

SPSS Advanced Models™ umfasst vor allem Verfahren, die in der fortgeschrittenen experimentellen und biomedizinischen Forschung Anwendung finden. Dies beinhaltet beispielsweise Prozeduren für allgemeine lineare Modelle (GLM), lineare gemischte Modelle, verallgemeinerte lineare Modelle (GZLM), verallgemeinerte Schätzungsgleichungen (GEE), Varianz-Komponentenanalyse, loglineare Analysen, versicherungsstatistische Sterbetafeln, die Überlebensanalyse nach Kaplan-Meyer sowie die grundlegende und erweiterte Cox-Regression.

SPSS Tables™ dient dem Erstellen einer großen Auswahl von Tabellenberichten in Präsentationsqualität. Mit dieser Option können beispielsweise komplexe Stub- und Banner-Tabellen erstellt und Daten von Mehrfachantworten angezeigt werden.

SPSS Trends™ bietet Funktionen zum Ausführen umfangreicher Prognosen sowie Zeitreihenanalysen mit Modellen für mehrfache Kurvenanpassung, mit Glättungsmodellen und Methoden zum Schätzen autoregressiver Funktionen.

SPSS Categories® bietet Funktionen zum Ausführen und Optimieren von Skalierungsprozeduren, u. a. Korrespondenzanalysen.

SPSS Conjoint™ bietet eine realistische Methode zum Messen, wie sich einzelne Produktmerkmale auf die Präferenzen von Konsumenten und Bürgern auswirken. Mit SPSS Conjoint können Sie einfach messen, welche Auswirkungen es hat, wenn einzelne Produktmerkmale im Kontext einer Gruppe von Produktmerkmalen gegeneinander abgewägt werden, genau wie Konsumenten dies bei Kaufentscheidungen tun.

SPSS Exact Tests™ berechnet exakte *P*-Werte für statistische Tests bei einer kleinen Anzahl oder sehr ungleichmäßig verteilten Stichproben, bei denen herkömmliche Tests nur ungenaue Ergebnisse liefern.

SPSS Missing Value Analysis™ dient zum Beschreiben von Mustern bei fehlenden Daten, zum Schätzen von Mittelwerten und anderen statistischen Größen sowie zum Ersetzen von Werten für fehlende Beobachtungen.

SPSS Maps™ bereitet geografisch verteilte Daten in Form von hochwertigen Karten mit Symbolen, Farben, Balkendiagrammen, Kreisdiagrammen und Themenkombinationen auf. So können Sie nicht nur das "Was", sondern auch das "Wo" zeigen.

SPSS Complex Samples™ ermöglicht Experten auf den Gebieten Umfragen, Marktforschung, Gesundheitswesen und Öffentliche Meinung sowie Sozialwissenschaftlern, die das Verfahren der Stichprobenumfrage verwenden, ihre Stichprobenpläne mit komplexen Stichproben in die Datenanalyse zu integrieren.

SPSS Classification Trees™ erstellt ein baumbasiertes Klassifizierungsmodell. Die Fälle werden in Gruppen klassifiziert oder es werden Werte für eine abhängige Variable (Zielvariable) auf der Grundlage der Werte von unabhängigen Variablen (Einflussvariablen) vorhergesagt. Die Prozedur umfasst Validierungswerkzeuge für die explorative und die bestätigende Klassifikationsanalyse.

Mit **SPSS Data Preparation™** erhalten Sie rasch eine visuelle Ansicht Ihrer Daten. Damit verfügen Sie über die Möglichkeit, Validierungsregeln anzuwenden, mit denen Sie ungültige Datenwerte identifizieren können. Sie können Regeln erstellen, mit denen Werte außerhalb des Bereichs, fehlende Werte oder leere Werte gekennzeichnet werden. Sie können außerdem Variablen speichern, mit denen individuelle Regelverletzungen sowie die Gesamtanzahl von Regelverletzungen je Fall aufgezeichnet werden. Im Lieferumfang des Programms befindet sich ein Satz von vordefinierten Regeln, die Sie kopieren und bearbeiten können.

Amos™ (**a**nalysis of **mo**ment **s**tructures) verwendet Modellierung von Strukturgleichungen, um konzeptuelle Modelle zu bestätigen und zu erklären, die auf Einstellungen, Wahrnehmungen und anderen Faktoren beruhen, die bestimmten Verhaltensweisen zugrunde liegen.

Zur Produktgruppe von SPSS gehören außerdem Anwendungen für Dateneingabe, Textanalyse, Klassifikation, neurale Netzwerke und Dienstleistungen für Unternehmen im Bereich der Prognose.

Installation

Zur Installation von SPSS Base System führen Sie den Lizenzautorisierungsassistenten mit dem Autorisierungscode aus, den Sie von SPSS erhalten haben. Weitere Informationen finden Sie in den Installationsanweisungen im Lieferumfang von SPSS Base System.

Kompatibilität

SPSS kann auf vielen Computersystemen ausgeführt werden. Mindestanforderungen an das System und Empfehlungen finden Sie in den Unterlagen, die mit Ihrem System geliefert werden.

Seriennummern

Die Seriennummer des Programms dient gleichzeitig als Identifikationsnummer bei SPSS. Sie benötigen diese Seriennummer, wenn Sie sich an SPSS wenden, um Informationen über Kundendienst, zu Zahlungen oder Aktualisierungen des Systems zu erhalten. Die Seriennummer wird mit dem Base-System ausgeliefert.

Kundendienst

Wenden Sie sich mit Fragen bezüglich der Lieferung oder Ihres Kundenkontos an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Halten Sie bitte stets Ihre Seriennummer bereit.

Ausbildungsseminare

SPSS bietet öffentliche und unternehmensinterne Seminare an. Alle Seminare beinhalten auch praktische Übungen. Seminare finden in größeren Städten regelmäßig statt. Wenn Sie weitere Informationen zu diesen Schulungen wünschen, wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden.

Technischer Support

Kunden von SPSS mit Wartungsvertrag können den Technischen Support in Anspruch nehmen. Kunden können sich an den Technischen Support wenden, wenn sie Hilfe bei der Arbeit mit SPSS oder bei der Installation in einer der unterstützten Hardware-Umgebungen benötigen. Informationen über den Technischen Support finden Sie auf der Website von SPSS unter *http://www.spss.com* oder wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Bei einem Anruf werden Sie nach Ihrem Namen, dem Namen Ihrer Organisation und Ihrer Seriennummer gefragt.

Weitere Veröffentlichungen

Weitere Exemplare von Produkthandbüchern können direkt bei SPSS Inc. bestellt werden. Besuchen Sie den SPSS Web Store unter *http://www.spss.com/estore* oder wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Wenden Sie sich bei telefonischen Bestellungen in den USA und Kanada unter 800-543-2185 direkt an SPSS Inc. Wenden Sie sich bei telefonischen Bestellungen außerhalb von Nordamerika an Ihr regionales SPSS-Büro, das Sie auf der SPPS-Website finden.

Das Handbuch *SPSS Statistical Procedures Companion* von Marija Norušis wurde von Prentice Hall veröffentlicht. Eine neue Fassung dieses Buchs mit Aktualisierungen für SPSS 16.0 ist geplant. Das Handbuch *SPSS Advanced Statistical Procedures Companion*, bei dem auch SPSS 16.0 berücksichtigt wird, erscheint demnächst. Das Handbuch *SPSS Guide to Data Analysis* für SPSS 16.0 wird ebenfalls derzeit erstellt. Ankündigungen für Veröffentlichungen,

die ausschließlich über Prentice Hall verfügbar sind, finden Sie auf der SPSS-Website unter *http://www.spss.com/estore* (wählen Sie Ihr Land aus und klicken Sie auf Books).

Kundenmeinungen

Ihre Meinung ist uns wichtig. Teilen Sie uns bitte Ihre Erfahrungen mit SPSS-Produkten mit. Insbesondere haben wir Interesse an neuen, interessanten Anwendungsgebieten von SPSS Base System. Senden Sie uns eine E-Mail an *suggest@spss.com* oder schreiben Sie an: SPSS Inc., Attn: Director of Product Planning, 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.

Über dieses Handbuch

In diesem Handbuch wird die grafische Benutzeroberfläche für die in SPSS Base System enthaltenen Prozeduren erläutert. Die Abbildungen der Dialogfelder stammen aus SPSS. Detaillierte Informationen zur Befehlssyntax für die Funktionen in SPSS Base System sind auf zwei Arten verfügbar: als Bestandteil des umfassenden Hilfesystems und als separates Dokument im PDF-Format im Handbuch *SPSS 16.0 Command Syntax Reference*, das auch über das Menü "Hilfe" verfügbar ist.

Kontakt zu SPSS

Wenn Sie in unseren Verteiler aufgenommen werden möchten, wenden Sie sich an eines unserer Büros, die Sie auf unserer Website unter *http://www.spss.com/worldwide* finden.

Inhalt

1 Übersicht 1

Neuerungen in Version 16.0 . . . 2
Windows . . . 4
Hauptfenster und aktives Fenster . . . 6
Statusleiste . . . 6
Dialogfelder . . . 7
Variablennamen und Variablenlabels in Listen von Dialogfeldern . . . 7
Ändern der Größe von Dialogfeldern . . . 8
Steuerelemente in Dialogfeldern . . . 8
Auswählen von Variablen . . . 9
Symbole für Datentyp, Messniveau und Variablenliste . . . 9
Aufrufen von Informationen zu Variablen in einem Dialogfeld . . . 10
Grundlegende Schritte bei der Datenanalyse . . . 10
Statistik-Assistent . . . 11
Weitere Informationen . . . 11

2 Aufrufen der Hilfe 12

Aufrufen der Hilfe zu ausgegebenen Begriffen . . . 13

3 Datendateien 15

Öffnen von Datendateien . . . 15
So öffnen Sie Datendateien: . . . 15
Datendateitypen . . . 16
Datei öffnen: Optionen . . . 17
Einlesen von Dateien aus Excel 95 oder nachfolgenden Versionen . . . 17
Einlesen von älteren Excel-Dateien und anderen Tabellenkalkulationsdateien . . . 17
Einlesen von dBASE-Dateien . . . 18
Einlesen von Stata-Dateien . . . 18
Einlesen von Datenbankdateien . . . 18
Text-Assistent . . . 33
Einlesen von Daten aus Dimensions . . . 42
Informationen zur Datei . . . 46

Speichern von Datendateien. 46
So speichern Sie geänderte Datendateien: 46
Speichern von Datendateien in externen Formaten 47
Speichern von Datendateien im Excel-Format 49
Speichern von Datendateien im SAS-Format 50
Speichern von Datendateien im Stata-Format 52
Speichern von Untergruppen von Variablen 53
Export in eine Datenbank 54
Export nach Dimensions 66
Schützen der ursprünglichen Daten 67
Virtuelle aktive Datei. 68
Erstellen eines Zwischenspeichers für Daten. 69

4 Modus für verteilte Analysen 72

Login beim SPSS-Server 72
Hinzufügen und Bearbeiten von Einstellungen für die Server-Anmeldung 73
So wählen Sie einen Server aus, wechseln den Server oder fügen einen neuen Server hinzu: . 74
Suche nach verfügbaren Servern 75
Öffnen von Datendateien auf einem Remote-Server 76
Dateizugriff im Modus für lokale und verteilte Analysen 76
Verfügbarkeit von Prozeduren im Modus für verteilte Analysen. 78
Absolute und relative Pfadangaben 78

5 Daten-Editor 80

Datenansicht 80
Variablenansicht. 81
So zeigen Sie die Attribute von Variablen an und legen diese fest: 82
Variablennamen 83
Messniveau einer Variablen 84
Variablentyp 84
Variablenlabels 87
Wertelabels. 87
Einfügen von Zeilenumbrüchen in Labels 88
Fehlende Werte 88
Spaltenbreite. 89
Variablenausrichtung. 89
Zuweisen von Variablenattributen zu mehreren Variablen 89
Benutzerdefinierte Variablenattribute 91

Anpassen der Variablenansicht . . . 95
Rechtschreibprüfung bei Variablen- und Wertelabels . . . 96
Eingeben von Daten . . . 96
So geben Sie numerische Daten ein: . . . 96
So geben Sie nichtnumerische Daten ein: . . . 97
So verwenden Sie Wertelabels bei der Dateneingabe: . . . 97
Einschränkungen für die Datenwerte im Daten-Editor . . . 97
Bearbeiten von Daten . . . 97
Ersetzen oder Ändern von Datenwerten . . . 98
Ausschneiden, Kopieren und Einfügen von Datenwerten . . . 98
Einfügen von neuen Fällen . . . 99
Einfügen von neuen Variablen . . . 99
So ändern Sie den Datentyp: . . . 100
Suchen von Fällen bzw. Variablen . . . 100
Suchen und Ersetzen von Daten- und Attributwerten . . . 101
Status für die Fallauswahl im Daten-Editor . . . 102
Optionen für die Anzeige im Daten-Editor . . . 103
Drucken aus dem Daten-Editor . . . 103
So drucken Sie den Inhalt des Daten-Editors: . . . 103

6 Arbeiten mit mehreren Datenquellen 105

Grundsätzlicher Umgang mit mehreren Datenquellen . . . 106
Arbeiten mit mehreren Daten-Sets in der Befehlssyntax . . . 107
Kopieren und Einfügen von Informationen zwischen Daten-Sets . . . 108
Umbenennen von Daten-Sets . . . 108
Unterdrücken der Anzeige mehrerer Daten-Sets . . . 109

7 Aufbereitung von Daten 110

Variableneigenschaften . . . 110
Definieren von Variableneigenschaften . . . 111
So definieren Sie Variableneigenschaften: . . . 111
Definieren von Wertelabels und anderen Variableneigenschaften . . . 112
Zuweisen des Messniveaus . . . 114
Benutzerdefinierte Variablenattribute . . . 115
Kopieren von Variableneigenschaften . . . 116
Mehrfachantworten-Sets . . . 117
Definieren von Mehrfachantworten-Sets . . . 117

Kopieren von Dateneigenschaften 120
So kopieren Sie Dateneigenschaften: 121
Auswählen von Quell- und Zielvariablen 122
Auswählen von Variableneigenschaften zum Kopieren 123
Kopieren der (Datei-)Eigenschaften eines Daten-Sets 124
Ergebnisse 127
Ermitteln doppelter Fälle 127
Visuelles Klassieren 131
So führen Sie die Klassierung von Variablen durch: 132
Klassieren von Variablen 132
Automatisches Erstellen von klassierten Kategorien 135
Kopieren von klassierten Kategorien 138
Benutzerdefinierte fehlende Werte in “Visuelles Klassieren” 139

8 Transformieren von Daten 140

Berechnen von Variablen 140
Variable berechnen: Falls Bedingung erfüllt ist 142
Variable berechnen: Typ und Label 142
Funktionen 143
Fehlende Werte in Funktionen. 143
Zufallszahlengeneratoren 144
Häufigkeiten von Werten in Fällen zählen 145
Werte in Fällen zählen: Welche Werte? 146
Häufigkeiten von Werten in Fällen zählen: Falls Bedingung erfüllt ist 146
Umkodieren von Werten 147
Umkodieren in dieselben Variablen. 147
Umkodieren in dieselben Variablen: Alte und neue Werte 148
Umkodieren in andere Variablen. 150
Umkodieren in andere Variablen: Alte und neue Werte 151
Rangfolge bilden. 152
Rangfolge bilden: Typen. 153
Rangfolge bilden: Bindungen. 154
Automatisch umkodieren 155
Assistent für Datum und Uhrzeit 158
Datums- und Zeitangaben in SPSS. 160
Erstellen einer Datums-/Zeitvariablen aus einer String-Variablen. 161
Erstellen einer Datums-/Zeitvariablen aus einem Variablen-Set 162
Addieren oder Subtrahieren von Werten zu bzw. von Datums-/Zeitvariablen 164
Extrahieren eines Teils einer Datums-/Zeitvariablen. 171

Datentransformationen für Zeitreihen 173
Datum definieren 174
Zeitreihen erstellen 175
Fehlende Werte ersetzen 178
Bewerten von Daten mit Vorhersagemodellen 180
Laden eines gespeicherten Modells 181
Anzeigen einer Liste der geladenen Modelle 183
Zusätzliche Funktionen bei der Befehlssyntax 184

9 Umgang mit Dateien und Dateitransformationen 185

Fälle sortieren 185
Variablen sortieren 186
Transponieren 188
Zusammenfügen von Datendateien 188
Fälle hinzufügen 189
Fälle hinzufügen: Umbenennen 192
Fälle hinzufügen: Informationen aus dem Datenlexikon 192
Zusammenfügen von mehr als zwei Datenquellen 192
Variablen hinzufügen 192
Variablen hinzufügen: Umbenennen 195
Zusammenfügen von mehr als zwei Datenquellen 195
Daten aggregieren 195
Daten aggregieren: Aggregierungsfunktion 198
Daten aggregieren: Variablenname und -label 198
Datei aufteilen 199
Fälle auswählen 200
Fälle auswählen: Falls 202
Fälle auswählen: Zufallsstichprobe 203
Fälle auswählen: Bereich 204
Fälle gewichten 204
Umstrukturieren von Daten 206
So strukturieren Sie Daten um: 206
Assistent für die Datenumstrukturierung: Auswählen des Typs 206
Assistent für die Datenumstrukturierung (Variablen zu Fällen): Anzahl von Variablengruppen 210
Assistent für die Datenumstrukturierung (Variablen zu Fällen): Auswählen Variablen 211
Assistent für die Datenumstrukturierung (Variablen zu Fällen): Erstellen von Indexvariablen . 213
Assistent für die Datenumstrukturierung (Variablen zu Fällen): Erstellen einer Indexvariablen 215
Assistent für die Datenumstrukturierung (Variablen zu Fällen): Erstellen mehrerer Indexvariablen 216

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Optionen 218
Assistent für die Datenumstrukturierung (Fälle zu Variablen): Auswählen von Variablen . . . 219
Assistent für die Datenumstrukturierung (Fälle zu Variablen): Sortieren von Daten 220
Assistent für die Datenumstrukturierung (Fälle zu Variablen): Optionen 221
Assistent für die Datenumstrukturierung: Fertig stellen 222

10 Arbeiten mit Ausgaben 224

Viewer 224
Ein- und Ausblenden von Ergebnissen 225
Verschieben, Löschen und Kopieren von Ausgaben 225
Ändern der anfänglichen Ausrichtung 226
Ändern der Ausrichtung von Ausgabeobjekten 226
Gliederung des Viewers 226
Einfügen von Objekten im Viewer 228
Suchen und Ersetzen von Informationen im Viewer 229
Kopieren von Ausgaben in andere Anwendungen 231
So kopieren Sie Ausgabeobjekte und fügen diese in eine andere Anwendung ein: 231
Ausgabe exportieren 232
Optionen für HTML, Word/RTF und Excel 234
PowerPoint-Optionen 235
PDF-Optionen 235
Text: Optionen 237
Optionen zum Exportieren von Diagrammen 238
Ausdrucken von Viewer-Dokumenten 240
So drucken Sie Ausgaben und Diagramme: 240
Seitenansicht 240
Seitenattribute: Kopf-/Fußzeile 241
Seitenattribute: Optionen 243
Speichern der Ausgabe 244
So speichern Sie ein Viewer-Dokument: 244

11 Pivot-Tabellen 245

Bearbeiten von Pivot-Tabellen 245
Aktivieren von Pivot-Tabellen 245
Pivotieren einer Tabelle 245
Ändern der Anzeigereihenfolge der Elemente innerhalb einer Dimension 246
Verschieben von Zeilen und Spalten innerhalb eines Dimensionselements 246
Vertauschen von Zeilen und Spalten 247

Gruppieren von Zeilen oder Spalten 247
Aufheben der Gruppierung von Zeilen oder Spalten 247
Drehen von Zeilen- und Spaltenbeschriftungen 248
Arbeiten mit Schichten 248
Erstellen und Anzeigen von Schichten 248
Gehe zu Kategorie in Schicht 250
Ein- und Ausblenden von Elementen 251
Ausblenden von Zeilen und Spalten in einer Tabelle 251
Anzeigen ausgeblendeter Zeilen und Spalten in einer Tabelle 251
Aus- und Einblenden von Dimensionsbeschriftungen 252
Aus- und Einblenden von Tabellentiteln 252
Tabellenvorlagen 252
So weisen Sie neue Tabellenvorlagen zu: 252
So bearbeiten oder erstellen Sie Tabellenvorlagen: 253
Tabelleneigenschaften 253
So ändern Sie die Eigenschaften von Pivot-Tabellen: 254
Tabelleneigenschaften: Allgemein 254
Tabelleneigenschaften: Fußnoten 255
Tabelleneigenschaften: Zellenformate 256
Tabelleneigenschaften: Rahmen 258
Tabelleneigenschaften: Drucken 259
Zelleneigenschaften 260
Schriftart und Hintergrund 261
Formatwert 261
Ausrichtung und Ränder 262
Fußnoten und Erklärungen 263
Hinzufügen von Fußnoten und Erklärungen 263
So können Sie eine Erklärung aus- bzw. einblenden: 264
So blenden Sie Fußnoten in Tabellen ein und aus: 264
Fußnotenzeichen 264
Neunummerierung von Fußnoten 265
Breite der Datenzellen 265
Ändern der Spaltenbreite 265
Anzeigen der ausgeblendeten Rahmen in einer Pivot-Tabelle: 265
Auswählen von Zeilen und Spalten in Pivot-Tabellen 266
Drucken von Pivot-Tabellen 267
Festlegen von Tabellenumbrüchen für breite und lange Tabellen 267
Erstellen eines Diagramms aus einer Pivot-Tabelle 268

12 Arbeiten mit der Befehlssyntax 269

Regeln für die Syntax . . . 269
Übernehmen der Befehlssyntax aus Dialogfeldern . . . 271
So übernehmen Sie die Befehlssyntax aus Dialogfeldern: . . . 271
Kopieren von Syntax aus dem Ausgabe-Log . . . 271
So kopieren Sie die Syntax aus dem Ausgabe-Log: . . . 272
So führen Sie Befehlssyntax aus: . . . 273
Unicode-Syntaxdateien . . . 273
Mehrere Execute-Befehle . . . 274

13 Häufigkeiten 275

Häufigkeiten: Statistik . . . 276
Häufigkeiten: Diagramme . . . 278
Häufigkeiten: Format . . . 279

14 Deskriptive Statistiken 280

Deskriptive Statistik: Optionen . . . 281
Zusätzliche Funktionen beim Befehl DESCRIPTIVES . . . 283

15 Explorative Datenanalyse 284

Explorative Datenanalyse: Statistik . . . 286
Explorative Datenanalyse: Diagramme . . . 287
Explorative Datenanalyse: Potenztransformationen . . . 288
Explorative Datenanalyse: Optionen . . . 288
Zusätzliche Funktionen beim Befehl EXAMINE . . . 289

16 Kreuztabellen 290

Kreuztabellenschichten . . . 292
Kreuztabellen: Gruppierte Balkendiagramme . . . 292
Kreuztabellen: Statistik . . . 292

Kreuztabellen: Zellen anzeigen 295
Kreuztabellen: Tabellenformat 296

17 Zusammenfassen 297

Zusammenfassen: Optionen 299
Zusammenfassung: Statistik 299

18 Mittelwerte 302

Mittelwerte: Optionen 304

19 OLAP-Würfel 307

OLAP-Würfel: Statistiken 309
OLAP-Würfel: Differenzen 311
OLAP-Würfel: Titel 312

20 T-Tests 313

T-Test bei unabhängigen Stichproben 313
T-Test bei unabhängigen Stichproben: Gruppen definieren 315
T-Tests bei unabhängigen Stichproben: Optionen 316
T-Test bei gepaarten Stichproben 316
T-Test bei gepaarten Stichproben: Optionen 317
T-Test bei einer Stichprobe 318
T-Test bei einer Stichprobe: Optionen 319
Zusätzliche Funktionen beim Befehl T-TEST 320

21 Einfaktorielle ANOVA 321

Einfaktorielle ANOVA: Kontraste 322
Einfaktorielle ANOVA: Post-Hoc-Mehrfachvergleiche 323

Einfaktorielle ANOVA: Optionen 326
Zusätzliche Funktionen beim Befehl ONEWAY 327

22 GLM - Univariat 328

GLM: Modell 330
Terme konstruieren 331
Quadratsumme 331
GLM: Kontraste 332
Kontrasttypen 333
GLM: Profilplots 334
GLM: Post-Hoc-Vergleiche 335
GLM: Speichern 337
GLM: Optionen 338
Zusätzliche Funktionen beim Befehl UNIANOVA 340

23 Bivariate Korrelationen 341

Bivariate Korrelationen: Optionen 343
Zusätzliche Funktionen bei den Befehlen CORRELATIONS und NONPAR CORR 343

24 Partielle Korrelationen 345

Partielle Korrelationen: Optionen 347
Zusätzliche Funktionen beim Befehl PARTIAL CORR 347

25 Distanzen 349

Unähnlichkeitsmaße für Distanzen 351
Ähnlichkeitsmaße für Distanzen 352
Zusätzliche Funktionen beim Befehl PROXIMITIES 353

26 Lineare Regression 354

Lineare Regression: Methode zur Auswahl von Variablen . . . 355
Lineare Regression: Bedingung aufstellen . . . 357
Lineare Regression: Diagramme . . . 357
Lineare Regression: Speichern von neuen Variablen . . . 359
Lineare Regression: Statistiken . . . 362
Lineare Regression: Optionen . . . 363
Zusätzliche Funktionen beim Befehl REGRESSION . . . 364

27 Ordinale Regression 365

Ordinale Regression: Optionen . . . 366
Ordinale Regression: Ausgabe . . . 368
Ordinale Regression: Kategorie . . . 369
    Terme konstruieren . . . 371
Ordinale Regression: Skala . . . 370
    Terme konstruieren . . . 371
Zusätzliche Funktionen beim Befehl PLUM . . . 371

28 Kurvenanpassung 372

Modelle für die Kurvenanpassung . . . 374
Kurvenanpassung: Speichern . . . 375

29 Regression mit partiellen kleinsten Quadraten 376

Modell . . . 378
Optionen . . . 379

30 Diskriminanzanalyse 381

Diskriminanzanalyse: Bereich definieren . . . 383
Diskriminanzanalyse: Fälle auswählen . . . 383

Diskriminanzanalyse: Statistik 384
Diskriminanzanalyse: Schrittweise Methode. 385
Diskirminanzanalyse: Klassifizieren 386
Diskriminanzanalyse: Speichern 388
Zusätzliche Funktionen beim Befehl DISCRIMINANT 388

31 Faktorenanalyse *389*

Faktorenanalyse: Fälle auswählen 390
Faktorenanalyse: Deskriptive Statistiken. 391
Faktorenanalyse: Extraktion 392
Faktorenanalyse: Rotation 394
Faktorenanalyse: Faktorwerte 395
Faktorenanalyse: Optionen 396
Zusätzliche Funktionen beim Befehl FACTOR. 396

32 Auswählen einer Prozedur zum Durchführen einer Clusteranalyse *397*

33 Two-Step-Clusteranalyse *399*

Two-Step-Clusteranalyse: Optionen 402
Two-Step-Clusteranalyse: Diagramme 404
Two-Step-Clusteranalyse: Ausgabe 405

34 Hierarchische Clusteranalyse *407*

Hierarchische Clusteranalyse: Methode 409
Hierarchische Clusteranalyse: Statistik 410
Hierarchische Clusteranalyse: Diagramme 410
Hierarchische Clusteranalyse: Neue Variablen 411
Zusätzliche Funktionen beim Befehl CLUSTER. 411

35 Clusterzentrenanalyse 412

Clusterzentrenanalyse: Effizienz 414
Clusterzentrenanalyse: Iterieren 414
Clusterzentrenanalyse: Neue Variablen 415
Clusterzentrenanalyse: Optionen 415
Zusätzliche Funktionen beim Befehl QUICK CLUSTER 416

36 Nichtparametrische Tests 417

Chi-Quadrat-Test 417
Chi-Quadrat-Test: erwarteter Bereich und erwartete Werte 419
Chi-Quadrat-Test: Optionen 419
Zusätzliche Funktionen beim Befehl NPAR TESTS (Chi-Quadrat-Test) 420
Test auf Binomialverteilung 420
Optionen für den Test auf Binomialverteilung 421
Zusätzliche Funktionen beim Befehl NPAR TESTS (Test auf Binomialverteilung) 422
Sequenzentest 422
Sequenzentest: Trennwert 423
Sequenzentest: Optionen 423
Zusätzliche Funktionen beim Befehl NPAR TESTS (Sequenzentest) 424
Kolmogorov-Smirnov-Test bei einer Stichprobe 424
K-S bei einer Stichprobe: Optionen 425
Zusätzliche Funktionen beim Befehl NPAR TESTS (Kolmogorov-Smirnov-Anpassungstest) 426
Tests bei zwei unabhängigen Stichproben 426
Typen von Tests bei zwei unabhängigen Stichproben 427
Zwei unabhängige Stichproben: Gruppen definieren 428
Tests bei zwei unabhängigen Stichproben – Optionen 428
Zusätzliche Funktionen beim Befehl NPAR TESTS (Tests bei zwei unabhängigen Stichproben) 429
Tests bei zwei verbundenen Stichproben 429
Typen von Tests bei zwei verbundenen Stichproben 430
Optionen für Tests bei zwei verbundenen Stichproben 431
Zusätzliche Funktionen beim Befehl NPAR TESTS (zwei verbundene Stichproben) 431
Tests bei mehreren unabhängigen Stichproben 431
Tests bei mehreren unabhängigen Stichproben: Welche Tests durchführen? 432
Tests bei mehreren unabhängigen Stichproben: Bereich definieren 433
Tests bei mehreren unabhängigen Stichproben: Optionen 433
Zusätzliche Funktionen beim Befehl NPAR TESTS (K unabhängige Stichproben) 434

Tests bei mehreren verbundenen Stichproben 434
Tests bei mehreren verbundenen Stichproben: Welche Tests durchführen?. 435
Tests bei mehreren verbundenen Stichproben: Statistiken. 435
Zusätzliche Funktionen beim Befehl NPAR TESTS (K verbundene Stichproben) 435

37 Analyse von Mehrfachantworten 436

Mehrfachantworten: Sets definieren 437
Mehrfachantworten: Häufigkeiten 438
Mehrfachantworten: Kreuztabellen 440
Mehrfachantworten: Kreuztabellen, Bereich definieren 442
Mehrfachantworten: Kreuztabellen, Optionen. 442
Zusätzliche Funktionen beim Befehl MULT RESPONSE 443

38 Ergebnisberichte 444

Bericht in Zeilen 444
So erstellen Sie eine Zusammenfassung: Bericht in Zeilen. 445
Datenspaltenformat/Break-Format in Berichten. 446
Bericht: Auswertungszeilen für/Endgültige Auswertungszeilen 447
Bericht: Break-Optionen 447
Bericht: Optionen 448
Bericht: Layout 449
Bericht: Titel 450
Bericht in Spalten. 450
So erstellen Sie eine Zusammenfassung: Bericht in Spalten 451
Datenspalten: Auswertungsfunktion 452
Auswertungsspalte für Gesamtergebnis 453
Format der Berichtsspalte 454
Bericht: Break-Optionen für Bericht in Spalten 454
Bericht: Optionen für Bericht in Spalten. 454
Bericht: Layout für Bericht in Spalten 455
Zusätzliche Funktionen beim Befehl REPORT 455

39 Reliabilitätsanalyse 456

Reliabilitätsanalyse: Statistik 458
Zusätzliche Funktionen beim Befehl RELIABILITY 460

40 Multidimensionale Skalierung 461

Multidimensionale Skalierung: Form der Daten . . . 463
Multidimensionale Skalierung: Distanzen aus Daten erstellen . . . 463
Multidimensionale Skalierung: Modell . . . 464
Multidimensionale Skalierung: Optionen . . . 465
Zusätzliche Funktionen beim Befehl ALSCAL . . . 466

41 Verhältnisstatistik 467

Verhältnisstatistik . . . 469

42 ROC-Kurven 471

ROC-Kurve: Optionen . . . 472

43 Übersicht über die Diagrammfunktion 474

Erstellen und Ändern von Diagrammen . . . 474
Erstellen von Diagrammen . . . 474
Bearbeiten von Diagrammen . . . 478
Optionen für die Diagrammdefinition . . . 481
Hinzufügen und Bearbeiten von Titeln und Fußnoten . . . 481
Festlegen von allgemeinen Optionen . . . 482

44 Extras 485

Variablenbeschreibungen . . . 485
Datendateikommentare . . . 486
Variablen-Sets . . . 486
Variablen-Sets definieren . . . 486
Variablen-Sets verwenden . . . 487
Umsortieren von Listen mit Zielvariablen . . . 489

45 Optionen 490

Optionen: Allgemein . . . 491
Optionen: Viewer . . . 493
Optionen: Daten . . . 494
Ändern der Standard-Variablenansicht . . . 496
Optionen: Währung . . . 496
So erstellen Sie Währungsformate: . . . 497
Optionen: Beschriftung der Ausgabe . . . 497
Diagrammoptionen . . . 499
Datenelement Farben . . . 500
Linien von Datenelementen . . . 500
Markierungen für Datenelemente . . . 501
Füllmuster für Datenelemente . . . 502
Pivottabellenoptionen . . . 502
Optionen für Datei-Speicherstellen . . . 504
Optionen: Skripte . . . 505

46 Anpassen von Menüs und Symbolleisten 508

Menü-Editor . . . 508
Anpassen von Symbolleisten . . . 509
Symbolleisten anzeigen . . . 509
So passen Sie Symbolleisten an: . . . 510
Symbolleiste: Eigenschaften . . . 511
Symbolleiste bearbeiten . . . 511
Neues Symbol erstellen . . . 512

47 Produktionsjobs 514

HTML-Optionen . . . 517
PowerPoint-Optionen . . . 518
PDF-Optionen . . . 518
Text-Optionen . . . 518
Laufzeitwerte . . . 519
Benutzerdefinierte Eingabeaufforderungen . . . 520

Ausführen von Produktionsjobs aus der Befehlszeile 521
Konvertieren von Produktionsmodus-Dateien 522

48 Ausgabeverwaltungssystem (OMS) 523

Ausgabeobjekttypen 526
Befehls-IDs und Tabellenuntertypen 527
Labels 528
OMS: Optionen 529
Protokollierung 534
Ausschließen der Ausgabeanzeige aus dem Viewer 534
Weiterleiten der Ausgabe an SPSS-Datendateien 535
Beispiel: Einzelne zweidimensionale Tabelle 535
Beispiel: Tabellen mit Schichten 536
Datendateien aus mehreren Tabellen 537
Steuern von Spaltenelementen zum Steuern von Variablen in der Datendatei 540
Variablennamen in Datendateien aus dem OMS 542
OXML-Tabellenstruktur 543
OMS-IDs 547
Kopieren von OMS-IDs aus Viewer-Gliederung 548

Anhang

Index 550

Kapitel
1

Übersicht

Mit SPSS verfügen Sie über ein leistungsfähiges System für statistische Analyse und Datenmanagement mit einer grafischen Benutzeroberfläche. Aussagekräftige Menüs und übersichtlich gestaltete Dialogfelder nehmen Ihnen einen großen Teil Ihrer Arbeit ab. Die meisten Aufgaben können einfach mit der Maus durchgeführt werden.

Neben der einfach zu bedienenden Benutzeroberfläche für die statistische Analyse finden Sie in SPSS die folgenden Hilfsmittel:

Daten-Editor. Der Daten-Editor ist ein vielseitiges System (ähnlich einer Tabellenkalkulation) für das Definieren, Eingeben, Bearbeiten und Anzeigen von Daten.

Viewer. Der Viewer erleichtert das Betrachten der Ergebnisse, das Ein- bzw. Ausblenden der Ausgaben, das Ändern der Anzeigereihenfolge der Ergebnisse und die Übertragung von Tabellen und Diagrammen in Präsentationsqualität von und aus anderen Anwendungen.

Mehrdimensionale Pivot-Tabellen. Mit mehrdimensionalen Pivot-Tabellen erwecken Sie Ihre Ergebnisse zum Leben. Sie können die Anordnung der Zeilen, Spalten und Schichten zum Auswerten Ihrer Tabellen ändern. So können Sie wichtige Ergebnisse hervorheben, die in "normalen" Berichten untergehen würden. Wenn Sie die Tabelle aufteilen, sodass immer nur eine Gruppe angezeigt wird, können Sie Gruppen leichter vergleichen.

Hochauflösende Grafiken. Hochauflösende und farbige Kreisdiagramme, Balkendiagramme, Histogramme, Streudiagramme, 3D-Grafiken und mehr sind als Standardfunktionen enthalten.

Datenbankzugriff. Sie können Informationen aus Datenbanken abrufen, indem Sie anstelle von komplizierten SQL-Abfragen den Datenbank-Assistenten verwenden.

Transformieren von Daten. Die Funktionen für das Transformieren von Daten erleichtern Ihnen die Vorbereitung Ihrer Daten für die Analyse. Unter anderem können Sie Daten problemlos in Untergruppen aufteilen und Kategorien kombinieren sowie Dateien hinzufügen, aggregieren, zusammenfügen, aufteilen und transponieren.

Online-Hilfe. Die ausführlichen Lernprogramme bieten Ihnen einen umfassenden Überblick über das Programm, kontextsensitive Hilfethemen in Dialogfeldern führen Sie durch bestimmte Aufgaben, und Popup-Definitionen in den Ergebnissen von Pivot-Tabellen erklären statistische Begriffe. Der Statistik-Assistent hilft Ihnen, geeignete Prozeduren zu finden, und die Fallstudien enthalten praktische Beispiele zum Verwenden von statistischen Prozeduren und Interpretieren der Ergebnisse.

Befehlssprache. Die meisten Aufgaben können einfach mit der Maus durchgeführt werden. Dennoch bietet SPSS auch eine leistungsfähige Befehlssprache, mit deren Hilfe Sie viele häufig durchzuführende Aufgaben speichern und automatisieren können. Die Befehlssprache bietet außerdem einige Funktionen, die nicht über die Menüs und Dialogfelder zur Verfügung stehen.

Eine vollständige Dokumentation zur Befehlssyntax findet sich im umfassenden Hilfesystem und als separates Dokument im PDF-Format in der *Command Syntax Reference*, die auch über das Menü “Hilfe” verfügbar ist.

Neuerungen in Version 16.0

Erweiterungen der Benutzeroberfläche. Auf der Benutzeroberfläche stehen nun folgende Erweiterungen zur Verfügung:

- Nun kann bei sämtlichen Dialogfeldern die Größe geändert werden. Durch die Verbreiterung eines Dialogfelds werden auch die Variablenlisten breiter, sodass Sie mehr von den Variablennamen bzw. beschreibenden Labels sehen können. Durch die Verlängerung eines Dialogfelds wird auch die Variablenliste länger, sodass Sie mehr Variablen anzeigen können, ohne einen Bildlauf durchführen zu müssen.
- Die Variablenauswahl mittels Ziehen und Ablegen wird nun in allen Dialogfeldern unterstützt.
- Die Anzeigereihenfolge und Anzeigeeigenschaften der Variablenlisten kann in allen Dialogfeldern im laufenden Programm aktualisiert werden. Sie können jederzeit die Sortierreihenfolge (alphabetisch, Dateireihenfolge, Messniveau) ändern bzw. zwischen der Anzeige der Variablennamen und der Variablenlabels umschalten. Für weitere Informationen siehe Variablennamen und Variablenlabels in Listen von Dialogfeldern auf S. 7.

Daten- und Ausgabeverwaltung. Folgende Erweiterungen für die Daten- und Ausgabeverwaltung stehen nun zur Verfügung:

- Lesen und Schreiben von Excel 2007-Dateien.
- Auswahl zwischen dem Arbeiten mit mehreren Daten-Sets oder jeweils einem einzigen Daten-Set. Für weitere Informationen siehe Optionen: Allgemein in Kapitel 45 auf S. 491.
- Suchen und Ersetzen von Informationen in Viewer-Dokumenten, einschließlich ausgeblendeter Objekte und Schichten in mehrdimensionalen Pivot-Tabellen. Für weitere Informationen siehe Suchen und Ersetzen von Informationen im Viewer in Kapitel 10 auf S. 229.
- Zuweisen fehlender Werte und Wertelabels zu beliebigen Stringvariablen, unabhängig von der definierten Stringlänge (zuvor auf Strings mit einer definierten Länge von maximal 8 Byte beschränkt).
- Neue zeichenbasierte Stringfunktionen.
- OMS-Unterstützung (Output Management System, Ausgabeverwaltungssystem) für Diagramme mit Viewer-Dateiformat (*.spv*) und VML-Format und Image Maps mit Popup-Diagramminformationen für HTML-Dokumente. Für weitere Informationen siehe Ausgabeverwaltungssystem (OMS) in Kapitel 48 auf S. 523.
- Anpassen der Variablenansicht im Daten-Editor. Sie können die Anzeigereihenfolge der Attributspalten ändern und festlegen, welche Attributspalten angezeigt werden sollen. Für weitere Informationen siehe Anpassen der Variablenansicht in Kapitel 5 auf S. 95.
- Sortieren der Variablen in der Arbeitsdatei alphabetisch oder nach Attributwerten (Wörterbuchwerten). Für weitere Informationen siehe Variablen sortieren in Kapitel 9 auf S. 186.

- Rechtschreibprüfung für Variablen- und Wertelabels in der Variablenansicht. Für weitere Informationen siehe Rechtschreibprüfung bei Variablen- und Wertelabels in Kapitel 5 auf S. 96.
- Ändern des grundlegenden Variablentyps (String, numerisch), Ändern der definierten Länge von Stringvariablen und automatische Festlegung der Länge von Stringvariablen auf den längsten beobachteten Wert für die betreffende Variable.
- Lesen und Schreiben von Daten- und Syntaxdateien in Unicode. Für weitere Informationen siehe Optionen: Allgemein in Kapitel 45 auf S. 491.
- Festlegen des Standardverzeichnisses zum Suchen nach und Speichern von Dateien. Für weitere Informationen siehe Optionen für Datei-Speicherstellen in Kapitel 45 auf S. 504.

Leistung. Bei Computern mit mehreren Prozessoren oder Prozessoren mit mehreren Cores steht nun für einige Prozeduren Multithreading für eine höhere Leistung zur Verfügung.

Erweiterungen bei Statistiken. Für Statistiken stehen nun folgende Erweiterungen zur Verfügung:

- Partielle kleinste Quadrate (Partial Least Squares, PLS) Ein Vorhersageverfahren, das eine Alternative zum Regressionsmodell der gewöhnlichen kleinsten Quadrate (Ordinary Least Squares, OLS), zur kanonischen Korrelation bzw. zur Modellierung von Strukturgleichungen darstellt und besonders nützlich ist, wenn die Einflussvariablen eine hohe Korrelation aufweisen oder wenn die Anzahl der Einflussvariablen die Anzahl der Fälle übersteigt. Für weitere Informationen siehe Regression mit partiellen kleinsten Quadraten in Kapitel 29 auf S. 376.
- Mehrschichtiges Perzeptron (MLP). Die MLP-Prozedur passt eine bestimmte Art von neuronalen Netzwerk an, ein so genanntes mehrschichtiges Perzeptron. Das mehrschichtige Perzeptron verwendet eine Feedforward-Architektur und kann mehrere verborgene Schichten enthalten. Das mehrschichtige Perzeptron ist hinsichtlich der Modelltypen, die es anpassen kann, sehr flexibel. Es ist eine der am häufigsten verwendeten Architekturen für neuronale Netze. Diese Prozedur ist in der neuen Option neuronales Netz verfügbar.
- Radiale Basisfunktion (RBF). Ein Netzwerk mit radialen Basisfunktionen (RBF) ist ein Feedforward-Netzwerk mit überwachtem Lernen und nur einer einzigen verborgenen Schicht, der so genannten radialen Basisfunktionsschicht. Wie beim Netzwerk mit mehrschichtigen Perzeptronen (MLP-Netzwerk) sind auch mit dem RPF-Netzwerk sowohl Vorhersagen als auch Klassifizierungen möglich. Es kann viel schneller sein als MLP, es ist jedoch nicht so flexibel hinsichtlich der Modelltypen, für die eine Anpassung möglich ist. Diese Prozedur ist in der neuen Option neuronales Netz verfügbar.
- "Verallgemeinerte lineare Modelle" unterstützt zahlreiche neue Funktionen, darunter ordinale multinomiale Verteilungen und Tweedie-Verteilungen, Maximum-Likelihood-Schätzung des negativen binomialen Hilfsparameters und Likelihood-Quotienten-Statistiken. Diese Prozedur ist in SPSS Advanced Models verfügbar.
- Die Cox-Regression bietet nun die Möglichkeit, Modellinformationen in eine XML-Datei (PMML) zu exportieren. Diese Prozedur ist in SPSS Advanced Models verfügbar.
- Cox-Regression für komplexe Stichproben. Wendet die Cox-Regression mit proportionalen Hazards auf Überlebenszeiten an, also auf die Zeitspanne vor dem Eintreten eines Ereignisses für Stichproben, die mit Methoden für komplexe Stichproben gezogen wurden. Diese Prozedur unterstützt stetige und kategoriale Einflussvariablen, die zeitabhängig sein können. Diese Prozedur stellt eine einfache Methode zur Berücksichtigung der Unterschiede in Untergruppen

sowie zur Analyse von Effekten eines Sets an Einflussvariablen dar. Die Prozedur schätzt Varianzen, indem sie den zur Auswahl der Stichprobe verwendeten Stichprobenplan berücksichtigt, einschließlich der Methoden mit gleichen Wahrscheinlichkeiten und der PPS-Methoden (PPS: probability proportional to size; Wahrscheinlichkeit proportional zum Umfang) und der Stichprobenprozeduren mit und ohne Zurücklegen (MZ bzw. OZ) Diese Prozedur ist in der Option "Complex Samples" (Komplexe Stichproben) verfügbar.

Programmierbarkeitserweiterung. Folgende Programmierbarkeitserweiterungen stehen zur Verfügung:

- R-Plugin. Sie können nun die Leistungsfähigkeit von SPSS mit der Möglichkeit, Ihre eigenen Statistikroutinen mit R zu schreiben, kombinieren. Dieses Plugin ist nur als Download von der Seite www.spss.com/devcentral (*http://www.spss.com/devcentral*) verfügbar.
- Verschachtelte `Begin Program-End Program`-Befehlsstrukturen.
- Möglichkeit zum Erstellen und Verwalten mehrerer Daten-Sets.

Befehlssyntax. Eine vollständige Liste der Ergänzungen und Änderungen zur Befehlssyntax finden Sie im Abschnitt *Release History* im Kapitel *Introduction* der *Command Syntax Reference* (im Hilfe-Menü verfügbar).

Nicht mehr unterstützte Funktionen

- Es gibt keinen separaten Diagrammeditor für "interaktive" Diagramme mehr. Diagramme, die über die früheren Dialogfelder für "interaktive" Diagramme und über die `IGRAPH`-Befehlssyntax erstellt wurden, werden nun im selben Format erstellt wie alle anderen Diagramme und im selben Diagrammeditor bearbeitet.
- Einige Funktionen die in früheren Dialogfeldern für "interaktive" Diagramme und die in der `IGRAPH`-Befehlssyntax angegeben wurden, sind nicht mehr verfügbar.
- Der Text-Viewer steht nicht mehr zur Verfügung.
- Viewer-Dateien, die in früheren Versionen von SPSS erstellt wurden (*.spo*-Dateien) können in SPSS 16.0 nicht geöffnet werden. Bei Windows-Betriebssystemen beinhaltet die Installations-CD einen Legacy-Viewer, den Sie installieren können, um Viewer-Dateien anzuzeigen und zu bearbeiten, die in früheren Versionen erstellt wurden.
- Die Option "Maps" (Karten) steht nicht mehr zur Verfügung.
- Es stehen keine Dialogfelder für die veralteten Prozeduren bei den Optionen "Trends" und "Tables" (Tabellen) mehr zur Verfügung. Bei "Trends" erstreckt sich dies auf folgende Befehle: `AREG`, `ARIMA` und `EXSMOOTH`. Bei "Tables" erstreckt sich dies auf den Befehl `TABLES`. Wenn Sie über eine Lizenz für eine dieser Optionen verfügen, die die veralteten Prozeduren beinhaltet, wird die Befehlssyntax für diese Befehle weiterhin unterstützt.

Windows

In SPSS gibt es verschiedene Arten von Fenstern:

Daten-Editor. Der Daten-Editor zeigt den Inhalt der Datendatei an. Im Daten-Editor können Sie neue Datendateien erstellen und vorhandene Datendateien bearbeiten. Wenn Sie mehr als eine Datendatei geöffnet haben, besitzt jede Datendatei ein separates Fenster im Daten-Editor.

Viewer. Alle statistischen Ergebnisse, Tabellen und Diagramme werden im Viewer angezeigt. Sie können die Ausgaben bearbeiten und zur späteren Verwendung speichern. Das Fenster des Viewers wird automatisch geöffnet, wenn Sie das erste Mal eine Prozedur aufrufen, die eine Ausgabe erzeugt.

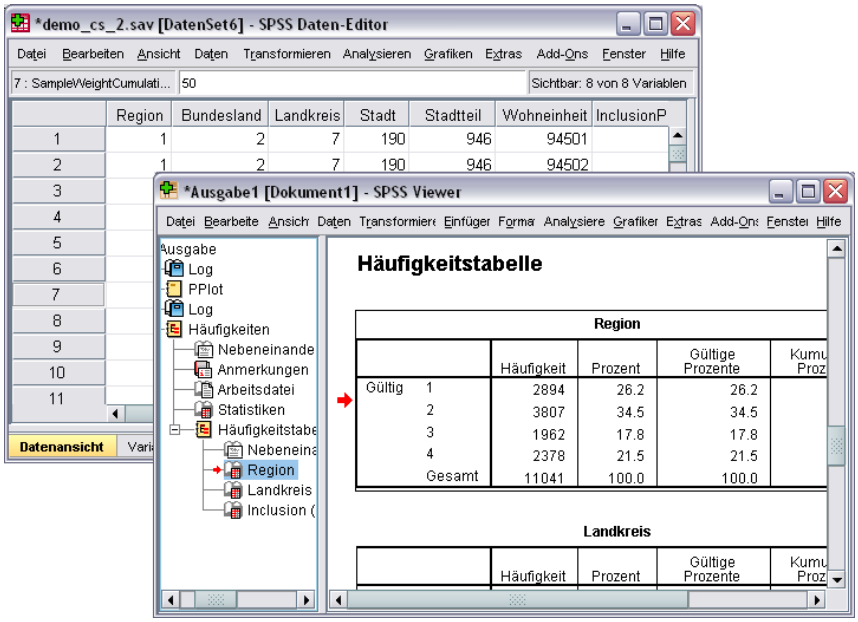
Pivot-Tabellen-Editor. Im Pivot-Tabellen-Editor verfügen Sie über vielseitige Möglichkeiten zur Bearbeitung von Ausgaben, die als Pivot-Tabellen angezeigt werden. Sie können Text bearbeiten, die Daten in Zeilen und Spalten austauschen, Farben hinzufügen, mehrdimensionale Tabellen erstellen und Ergebnisse ein- bzw. ausblenden.

Diagramm-Editor. In Diagrammfenstern können hochauflösende Diagramme und Grafiken bearbeitet werden. Sie können die Farben ändern, andere Schriftarten oder -größen auswählen, horizontale und vertikale Achsen vertauschen, 3D-Streudiagramme rotieren und den Diagrammtyp ändern.

Textausgabe-Editor. Textausgaben, die nicht in Pivot-Tabellen angezeigt werden, können im Textausgabe-Editor bearbeitet werden. Sie können die Ausgabe bearbeiten und die Schriftmerkmale ändern (Schriftart, Stil, Farbe, Größe).

Syntax-Editor. Sie können die in einem Dialogfeld getroffene Auswahl auch als Befehlssyntax direkt in ein Syntax-Fenster einfügen. Dort werden die Befehle als Befehlssyntax angezeigt. Sie können die Befehlssyntax dann bearbeiten und so die Funktionen nutzen, auf die Sie nicht über Dialogfelder zugreifen können. Diese Befehle können zur Nutzung in späteren Sitzungen in einer Datei gespeichert werden.

Abbildung 1-1
Daten-Editor und Viewer



Hauptfenster und aktives Fenster

Wenn Sie mehr als ein Viewer-Fenster geöffnet haben, werden Ausgaben an das **Hauptfenster** des Viewers geleitet. Wenn Sie mehr als ein Fenster für den Syntax-Editor geöffnet haben, wird Syntax in das Hauptfenster des Syntax-Editors eingefügt. Die Hauptfenster weisen ein Pluszeichen in der Titelleiste auf. Sie können die Hauptfenster jederzeit wechseln.

Das Hauptfenster darf nicht mit dem **aktiven** Fenster verwechselt werden. Das aktive Fenster ist das aktuell ausgewählte Fenster. Wenn sich mehrere Fenster überlappen, wird das aktive Fenster im Vordergrund angezeigt. Wenn Sie ein neues Fenster öffnen, wird dieses automatisch sowohl das aktive Fenster als auch das Hauptfenster.

Wechseln des Hauptfensters

- Aktivieren Sie das Fenster, welches das Hauptfenster werden soll. Klicken Sie dazu auf eine beliebige Stelle im Fenster.
- Klicken Sie auf die Schaltfläche “Hauptfenster” in der Symbolleiste (das Symbol mit dem Pluszeichen).

  oder

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Extras
  Hauptfenster

Anmerkung: Im Daten-Editor wird durch das aktive Daten-Editor-Fenster festgelegt, welches Daten-Set in Berechnungen oder Analysen verwendet wird. Es gibt kein “Hauptfenster” im Daten-Editor. Für weitere Informationen siehe Grundsätzlicher Umgang mit mehreren Datenquellen in Kapitel 6 auf S. 106.

Statusleiste

Die Statusleiste am unteren Rand jedes Fensters von SPSS enthält die folgenden Informationen:

Befehlsstatus. Bei jeder ausgeführten Prozedur und jedem ausgeführten Befehl wird angezeigt, wie viele Fälle bereits verarbeitet wurden. Bei statistischen Prozeduren mit iterativer Verarbeitung wird die Anzahl der Iterationen angezeigt.

Filterstatus. Wenn Sie eine zufällige Stichprobe oder eine Teilmenge der Fälle zum Analysieren ausgewählt haben, zeigt Ihnen die Meldung Filter an, dass die Fälle gefiltert und somit nicht alle Fälle in der Datendatei bei der Analyse berücksichtigt werden.

Gewichtungsstatus. Die Meldung Gewichtung an zeigt an, dass die Fälle mit einer Gewichtungsvariablen analysiert werden.

Aufspaltungsstatus der Datei. Die Meldung Datei aufteilen an bedeutet, dass die Datendatei zur Analyse anhand der Werte von einer oder mehr Gruppenvariablen in verschiedene Gruppen aufgeteilt wurde.

Dialogfelder

Wenn Sie eine Option aus einem Menü auswählen, wird in den meisten Fällen ein Dialogfeld geöffnet. Sie verwenden die Dialogfelder zum Auswählen von Variablen und Optionen für die Analyse.

Dialogfelder für statistische Prozeduren und Diagramme weisen in der Regel folgende zwei Hauptkomponenten auf:

Liste der Quellvariablen. Eine Variablenliste in der Arbeitsdatei. In der Quellliste werden nur die Typen von Variablen angezeigt, die für die ausgewählte Prozedur zulässig sind. Kurze und lange String-Variablen können in vielen Prozeduren nur eingeschränkt verwendet werden.

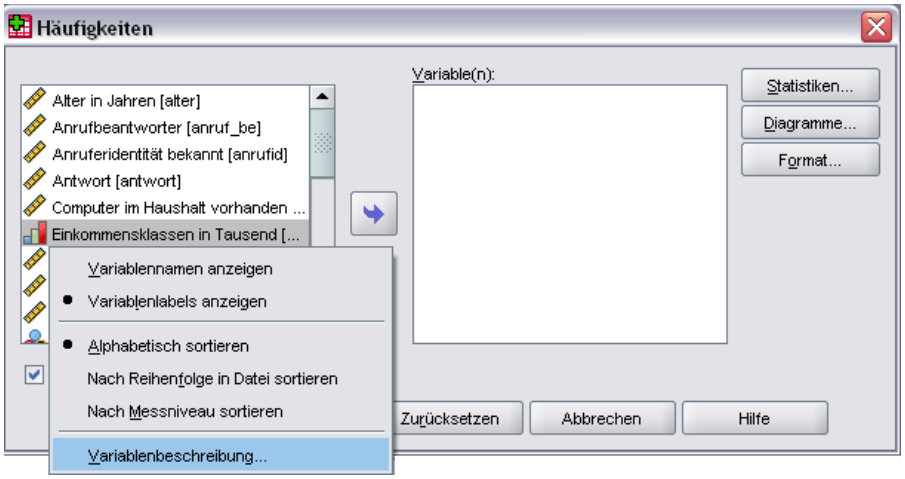
Zielvariablenliste(n). Hierbei handelt es sich um eine oder mehrere Listen mit den Variablen, die Sie zur Analyse ausgewählt haben, also beispielsweise die Liste mit den abhängigen und die Liste mit den unabhängigen Variablen.

Variablennamen und Variablenlabels in Listen von Dialogfeldern

In den Listen von Dialogfeldern können Sie entweder Variablennamen oder Variablenlabels anzeigen lassen. Außerdem können Sie die Sortierreihenfolge der Variablen in den Listen der Quellvariablen festlegen.

- Um die standardmäßig geltenden Anzeigeattribute der Variablen in Quelllisten festzulegen, wählen Sie im Menü “Bearbeiten” den Eintrag Optionen. Für weitere Informationen siehe Optionen: Allgemein in Kapitel 45 auf S. 491.
- Um die Anzeigeattribute für die Quellvariablenliste innerhalb eines Dialogfelds zu ändern, klicken Sie mit der rechten Maustaste auf eine beliebige Variable in der Quellliste und wählen Sie die gewünschten Anzeigeattribute im Kontextmenü aus. Sie können entweder Variablennamen oder Variablenlabels anzeigen lassen (bei allen Variablen, für die keine Labels definiert sind, werden die Namen angezeigt), und Sie können die Quellliste nach Dateireihenfolge, alphabetischer Reihenfolge oder nach Messniveau sortieren. Weitere Informationen zum Messniveau finden Sie unter Symbole für Datentyp, Messniveau und Variablenliste auf S. 9.

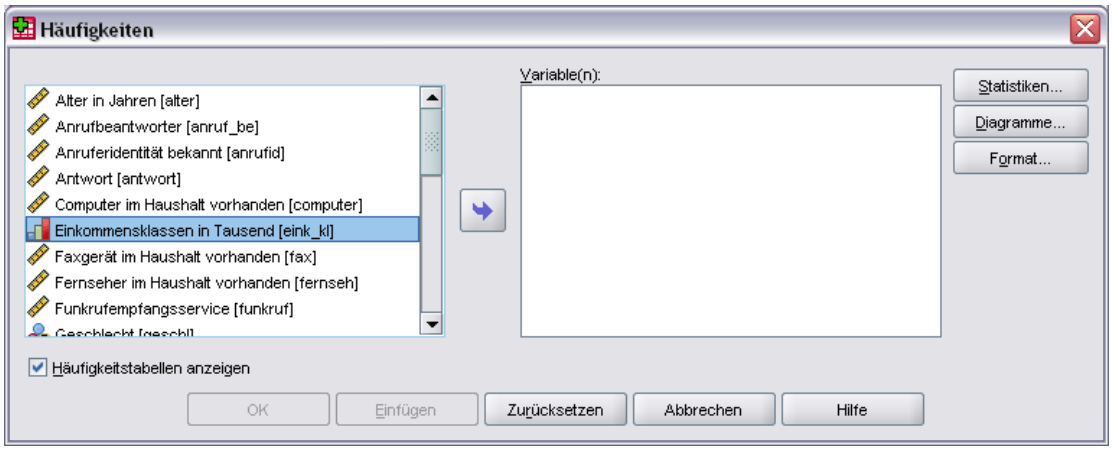
Abbildung 1-2
In Dialogfeldern angezeigte Variablenlabels



Ändern der Größe von Dialogfeldern

Sie können die Größe von Dialogfeldern ebenso ändern wie die von Fenstern, einfach indem Sie auf die äußeren Ränder oder Ecken klicken und daran ziehen. Wenn Sie beispielsweise das Dialogfeld verbreitern, werden auch die Variablenlisten breiter.

Abbildung 1-3
Dialogfeld nach der Größenänderung



Steuerelemente in Dialogfeldern

Die meisten Dialogfelder enthalten fünf Steuerelemente:

OK. Hiermit führen Sie die Prozedur aus. Klicken Sie nach dem Auswählen von Variablen und dem Festlegen von zusätzlichen Angaben auf OK, um die Prozedur auszuführen und das Dialogfeld zu schließen.

Einfügen. Hiermit erzeugen Sie aus den Einstellungen im Dialogfeld Befehlssyntax und fügen diese Syntax in ein Syntax-Fenster ein. Sie können die Befehle um zusätzliche Funktionen erweitern, auf die Sie sonst nicht über die Dialogfelder zugreifen können.

Zurücksetzen. Hiermit heben Sie die Auswahl von Variablen in den Listen der ausgewählten Variablen auf und setzen alle Einstellungen im aktuellen Dialogfeld und in allen untergeordneten Dialogfeldern auf die Standardeinstellungen zurück.

Abbrechen. Hiermit verwerfen Sie alle Änderungen, die seit dem letzten Öffnen an den Einstellungen im Dialogfeld vorgenommen wurden, und schließen das Dialogfeld. Innerhalb einer Sitzung bleiben die Einstellungen in einem Dialogfeld bestehen. Die Einstellungen in einem Dialogfeld werden beibehalten, bis Sie diese überschreiben.

Hilfe. Stellt kontextsensitive Hilfe bereit. Hiermit wechseln Sie zu einem Hilfefenster mit Informationen zum aktuellen Dialogfeld.

Auswählen von Variablen

Um eine einzelne Variablen auszuwählen, markieren Sie sie einfach in der Liste der Quellvariablen und verschieben Sie sie mittels Ziehen und Ablegen in die Liste der Zielvariablen. Sie können Variablen auch mithilfe der Pfeilschaltflächen aus der Quellliste in die Ziellisten verschieben. Wenn nur eine Liste für Zielvariablen vorhanden ist, können Sie auf einzelne Variablen doppelklicken, um diese aus der Liste der Quellvariablen in die Liste der Zielvariablen zu verschieben.

Sie können auch mehrere Variablen gleichzeitig auswählen:

- Um mehrere Variablen auszuwählen, die nacheinander in der Variablenliste stehen, klicken Sie auf die erste Variable, halten Sie die Umschalttaste gedrückt und klicken Sie anschließend auf die letzte Variable in der Gruppe.
- Um mehrere Variablen auszuwählen, die nicht nacheinander in der Variablenliste stehen, klicken Sie auf die erste Variable und halten Sie beim Klicken auf die weiteren Variablen die Strg-Taste gedrückt. (Macintosh: Halten Sie beim Klicken die Befehlstaste gedrückt.)

Symbole für Datentyp, Messniveau und Variablenliste

Die Symbole, die neben den Variablen in Dialogfeldern angezeigt werden, liefern Informationen über den Variablentyp und das Messniveau.

Messniveau	**Datentyp**			
	Numerisch	String	Datum	Zeit
Metrisch		entfällt		
Ordinal				
Nominal				

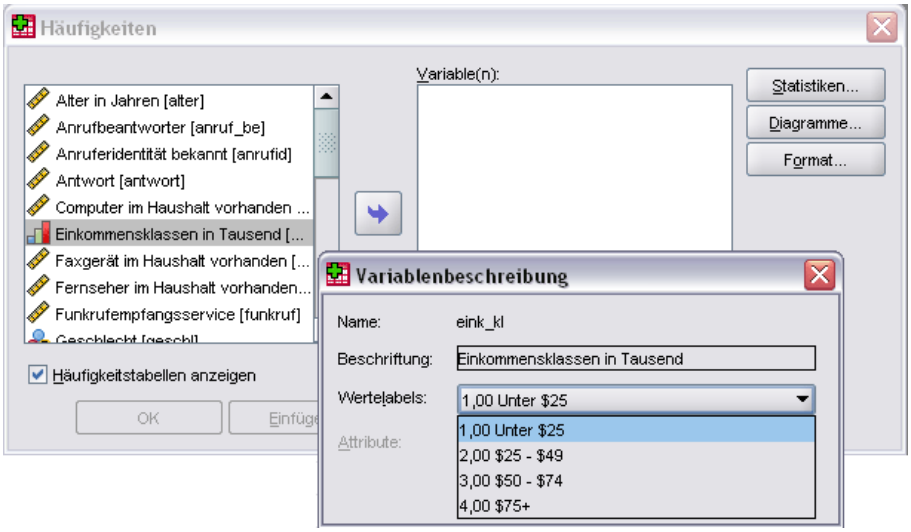
- Weitere Informationen zum Messniveau finden Sie unter Messniveau einer Variablen auf S. 84.
- Weitere Informationen zu den Datentypen “numerisch”, “String”, “Datum” und “Uhrzeit” finden Sie unter Variablentyp auf S. 84.

Aufrufen von Informationen zu Variablen in einem Dialogfeld

▶ Klicken Sie mit der rechten Maustaste auf eine Variable in der Liste der Quell- oder Zielvariablen.

▶ Wählen Sie Variablenbeschreibung aus.

Abbildung 1-4
Variablenbeschreibung



Grundlegende Schritte bei der Datenanalyse

Das Analysieren von Daten in SPSS ist sehr einfach. Sie müssen nur die folgenden Schritte ausführen:

Übertragen Ihrer Daten in SPSS. Sie können eine bereits gespeicherte SPSS-Datendatei öffnen; eine Tabellenblatt-, Datenbank- oder Text-Datendatei einlesen oder die Daten direkt im Daten-Editor eingeben.

Auswählen einer Prozedur. Wählen Sie eine Prozedur aus den Menüs aus, um Statistiken zu berechnen oder ein Diagramm zu erstellen.

Auswählen der Variablen für die Analyse. Die Variablen in der Datendatei werden in einem Dialogfeld für die Prozedur angezeigt.

Führen Sie die Prozedur aus, und betrachten Sie die Ergebnisse. Die Ergebnisse werden im Viewer angezeigt.

Statistik-Assistent

Wenn Ihnen SPSS oder die verfügbaren statistischen Prozeduren nicht vertraut sind, kann Ihnen der Statistik-Assistent den Einstieg erleichtern. Einfache Fragen in einer verständlichen Sprache und visuelle Beispiele helfen Ihnen bei der Auswahl der grundlegenden statistischen Funktionen und Diagrammfunktionen, die sich am besten für Ihre Daten eignen.

Zum Aufrufen des Statistik-Assistenten wählen Sie die folgenden Befehle aus den Menüs eines beliebigen Fensters in SPSS aus:

Hilfe
  Statistik-Assistent

Der Statistik-Assistent enthält nur eine ausgewählte Teilmenge der Prozeduren aus dem Base-System. Er dient der allgemeinen Unterstützung bei vielen grundlegenden und häufig verwendeten statistischen Verfahren.

Weitere Informationen

Eine umfassende Übersicht über die Grundlagen finden Sie im Online-Lernprogramm. Wählen Sie folgende Befehle aus den Menüs eines beliebigen Fensters von SPSS aus:

Hilfe
  Lernprogramm

Kapitel
2

Aufrufen der Hilfe

Hilfestellung ist auf verschiedene Weise verfügbar:

Menü "Hilfe". Das Menü "Hilfe" in den meisten SPSS-Fenstern führt zum Haupthilfesystem, außerdem zu Lernprogrammen und technischen Referenzen.

- **Themen.** Hiermit können Sie auf die Registerkarten "Inhalt", "Index" und "Suchen" zugreifen. Verwenden Sie diese Registerkarten bei der Suche nach bestimmten Hilfethemen.
- **Lernprogramm.** Illustrierte, schrittweise Anleitungen für die Verwendung zahlreicher Grundfunktionen in SPSS. Es ist nicht notwendig, das gesamte Lernprogramm vom Anfang bis zum Ende durchzuarbeiten. Sie können die gewünschten Themen direkt auswählen, nach Wunsch zwischen den Themen wechseln, die Themen in beliebiger Reihenfolge abrufen oder auch bestimmte Themen über den Index oder das Inhaltsverzeichnis suchen.
- **Fallstudien.** Praktische Beispiele für die Erstellung verschiedener Arten von statistischen Analysen und für die Interpretation der Ergebnisse. Die in den Beispielen verwendeten Datendateien werden auch bereitgestellt. Sie können also die Beispiele durcharbeiten, um zu verfolgen, wie die Ergebnisse zustande kommen. Sie können die gewünschten Prozeduren im Inhaltsverzeichnis auswählen oder nach relevanten Themen im Index suchen.
- **Statistik-Assistent.** Dieser Assistent unterstützt Sie bei der Suche nach der Prozedur, die Sie verwenden möchten. Nachdem Sie Ihre Auswahl getroffen haben, öffnet der Statistik-Assistent das Dialogfeld für die Statistik-, Berichts- oder Diagrammprozedur, das die ausgewählten Kriterien erfüllt. Der Statistik-Assistent bietet Zugang zu den meisten Statistik- und Berichtsprozeduren sowie auf einen Großteil der Diagrammprozeduren im Base-System.
- **Befehlssyntax-Referenz (Command Syntax Reference).** Detaillierte Informationen zur Befehlssyntax sind auf zwei Arten verfügbar: als Bestandteil des umfassenden Hilfesystems und als separates Dokument im PDF-Format im Handbuch *SPSS Command Syntax Reference*, das auch über das Menü "Hilfe" verfügbar ist.
- **Statistische Algorithmen.** Die für die meisten statistischen Prozeduren verwendeten Algorithmen sind in zwei Formaten verfügbar: als Bestandteil der umfassenden Hilfesystems und als separates Dokument im PDF-Format im Handbuch *SPSS Algorithms*, das auf der Handbuch-CD zur Verfügung steht. Links zu spezifischen Algorithmen im Hilfesystem erhalten Sie, wenn Sie im Hilfe-Menü die Option Algorithms auswählen.

Kontextsensitive Hilfe. An zahlreichen Stellen der Benutzeroberfläche können Sie kontextsensitive Hilfe abrufen.

- **Schaltflächen für Hilfe in Dialogfeldern.** Die meisten Dialogfelder verfügen über die Schaltfläche "Hilfe", mit der Sie das entsprechende Hilfethema für das Dialogfeld direkt aufrufen können. In diesem Hilfethema finden Sie allgemeine Informationen und Verknüpfungen zu verwandten Themen.

- **Hilfe zu Pivot-Tabellen über das Kontextmenü.** Wenn Sie mit der rechten Maustaste auf Begriffe einer im Viewer aktivierten Pivot-Tabelle klicken und dann Direkthilfe aus dem Kontextmenü auswählen, erhalten Sie eine Definition dieser Begriffe.
- **Befehlssyntax.** Zeigen Sie in einem Befehlssyntaxfenster auf eine beliebige Position innerhalb eines Syntaxblocks für einen Befehl und drücken Sie F1 auf der Tastatur. Das vollständige Befehlssyntaxdiagramm für diesen Befehl wird eingeblendet. Die vollständige Dokumentation für die Befehlssynatx ist über die Verknüpfungen in den Listen der verwandten Themen und auf der Registerkarte "Inhalt" der Hilfe verfügbar.

Sonstige Ressourcen

Website des technischen Supports. Antworten auf viele häufig auftretende Probleme finden Sie unter *http://support.spss.com*. (Für die Website des technischen Supports benötigen Sie eine Anmelde-ID und ein Passwort. Weitere Informationen zum Anfordern einer ID und eines Paßworts finden Sie unter der genannten URL.)

SPSS Developer Central. Developer Central bietet Ressourcen für SPSS-Benutzer und SPSS-Anwendungsentwickler auf allen Niveaus. Hier können Sie Dienstprogramme, Grafikbeispiele, neue Statistikmodule und Artikel zur SPSS-Technologie herunterladen. In den Foren können Sie mit SPSS und der SPSS-Benutzergemeinschaft in Kontakt treten. Besuchen Sie SPSS Developer Central unter *http://www.spss.com/devcentral*.

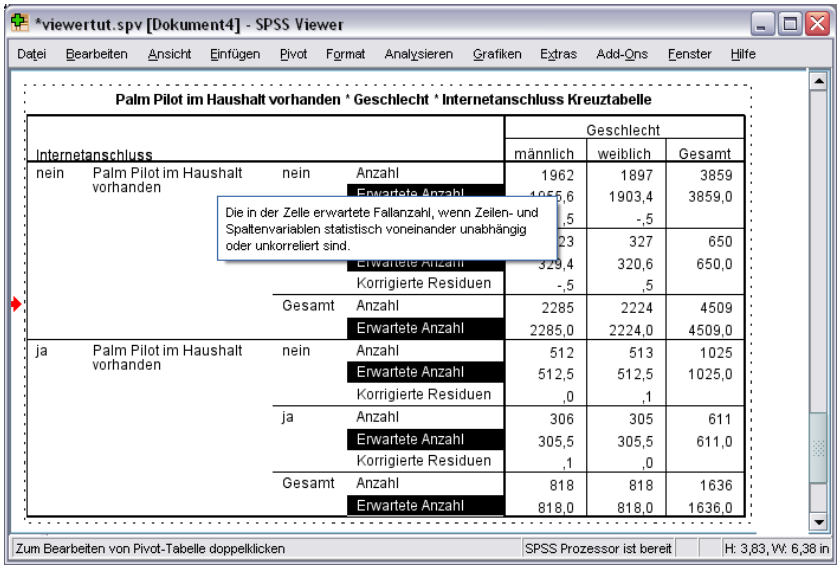
Aufrufen der Hilfe zu ausgegebenen Begriffen

So können Sie eine Definition für einen Begriff in einer Pivot-Tabellen-Ausgabe im Viewer anzeigen:

- ▶ Doppelklicken Sie auf die Pivot-Tabelle, um diese zu aktivieren.
- ▶ Klicken Sie mit der rechten Maustaste auf den Term, zu dem Sie Erklärungen benötigen.
- ▶ Wählen Sie Direkthilfe aus dem Kontextmenü aus.

  In einem Popup-Fenster wird eine Definition des Terms angezeigt.

Abbildung 2-1
Glossar/Hilfe zu aktiver Pivot-Tabelle über die rechte Maustaste



Internetanschluss				Geschlecht männlich	Geschlecht weiblich	Geschlecht Gesamt
nein	Palm Pilot im Haushalt vorhanden	nein	Anzahl	1962	1897	3859
			Erwartete Anzahl	1955,6	1903,4	3859,0
			Korrigierte Residuen	,5	-,5	
		ja	Anzahl	323	327	650
			Erwartete Anzahl	329,4	320,6	650,0
			Korrigierte Residuen	-,5	,5	
		Gesamt	Anzahl	2285	2224	4509
			Erwartete Anzahl	2285,0	2224,0	4509,0
ja	Palm Pilot im Haushalt vorhanden	nein	Anzahl	512	513	1025
			Erwartete Anzahl	512,5	512,5	1025,0
			Korrigierte Residuen	,0	,1	
		ja	Anzahl	306	305	611
			Erwartete Anzahl	305,5	305,5	611,0
			Korrigierte Residuen	,1	,0	
		Gesamt	Anzahl	818	818	1636
			Erwartete Anzahl	818,0	818,0	1636,0

Kapitel
3

Datendateien

Datendateien können in einer Reihe von verschiedenen Formaten vorliegen. SPSS kann mit vielen dieser Formaten arbeiten, unter anderem mit den folgenden:

- Tabellenkalkulationsblätter aus Excel und Lotus
- Datenbanktabellen aus vielen Datenbankquellen, einschließlich Oracle, SQL-Server, Access, DBASE und andere
- Mit Tabulatoren als Trennzeichen versehene und andere Typen von Textdateien
- Datendateien im SPSS-Format, die unter anderen Betriebssystemen erstellt wurden
- SYSTAT-Datendateien
- SAS-Datendateien
- Stata-Datendateien

Öffnen von Datendateien

Neben den im SPSS-Format gespeicherten Dateien lassen sich die Dateien von Excel, SAS und Stata sowie mit Tabulatoren als Trennzeichen versehene und andere Dateien öffnen, ohne diese Dateien in ein Zwischenformat umzuwandeln oder Datendefinitionseingaben vorzunehmen.

- Durch das Öffnen einer Datendatei wird diese zur Arbeitsdatei. Wenn Sie bereits eine oder mehrere Datendateien geöffnet wurden, bleiben diese geöffnet und für die anschließende Verwendung in der Sitzung verfügbar. Durch Klicken auf eine beliebige Stelle einer geöffneten Datendatei im Daten-Editor macht diese zur Arbeitsdatei. Für weitere Informationen siehe Arbeiten mit mehreren Datenquellen in Kapitel 6 auf S. 105.
- Im Modus für verteilte Analysen, bei dem ein Remote-Server zum Verarbeiten von Befehlen und Ausführen von Prozeduren verwendet wird, sind nur die Datendateien, Ordner und Laufwerke verfügbar, auf die der Remote-Server zugreifen kann. Der Name des aktuellen Servers wird im oberen Teil des Dialogfelds angezeigt. Sie können nur auf Datendateien auf dem lokalen Computer zugreifen, wenn Sie das Laufwerk oder die Ordner mit den Datendateien für den gemeinsamen Zugriff freigeben. Für weitere Informationen siehe Modus für verteilte Analysen in Kapitel 4 auf S. 72.

So öffnen Sie Datendateien:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Öffnen
    Daten...

- ▶ Wählen Sie im Dialogfeld “Daten öffnen” die zu öffnende Datei aus.
- ▶ Klicken Sie auf Öffnen.

Die folgenden Optionen sind verfügbar:

- Als Länge der einzelnen Stringvariablen kann automatisch der längste beobachtete Wert für die betreffende Variable festgelegt werden (mit der Option String-Längen anhand beobachteter Werte minimieren). Diese Funktion ist insbesondere beim Lesen von Codepage-Datendateien im Unicode-Modus nützlich. Für weitere Informationen siehe Optionen: Allgemein in Kapitel 45 auf S. 491.
- Aus der ersten Zeile von Tabellenkalkulationsdateien können die Namen von Variablen eingelesen werden.
- Bei Tabellenkalkulationsdateien kann der Zellenbereich angegeben werden, der eingelesen werden soll.
- Es kann angegeben werden, welches Arbeitsblatt aus der Excel-Datei gelesen werden soll (Excel 95 oder Nachfolgeversionen).

Informationen zum Einlesen von Daten aus Datenbanken finden Sie unter Einlesen von Datenbankdateien auf S. 18. Informationen zum Einlesen von Daten aus Textdatendateien finden Sie unter Text-Assistent auf S. 33.

Datendateitypen

SPSS. Hiermit werden Datendateien geöffnet, die im SPSS-Format sowie mit dem DOS-Produkt SPSS/PC+ gespeichert wurden.

SPSS/PC+. Hiermit werden Datendateien von SPSS/PC+ geöffnet.

SYSTAT. Hiermit werden Datendateien von SYSTAT geöffnet.

SPSS Portable. Hiermit werden Datendateien geöffnet, die im portablen Format gespeichert wurden. Das Speichern einer Datei im portablen Format erfordert erheblich mehr Zeit als das Speichern der Datei im SPSS-Format.

Excel. Hiermit werden Excel-Dateien geöffnet.

Lotus 1-2-3. Hiermit werden Datendateien geöffnet, die im Format von Lotus 1-2-3, den Versionen 3.0, 2.0 oder der Version 1A gespeichert wurden.

SYLK. Hiermit werden die im SYLK-Format gespeicherten Datendateien geöffnet. Dieses Format wird bei manchen Tabellenkalkulationsprogrammen eingesetzt.

dBASE. Hiermit werden dBASE-Dateien im Format von dBASE IV, dBASE III oder III PLUS sowie dBASE II geöffnet. Jeder Fall ist eine Zeile im Datensatz. Beim Speichern einer Datei in diesem Format gehen die Bezeichnungen von Variablen, die Wertelabels und die Angaben zu fehlenden Werten verloren.

SAS. SAS-Versionen 6–9 und Dateien im SAS-Transportformat.

Stata. Stata Versionen 4–8.

Datei öffnen: Optionen

Variablennamen lesen. Bei Tabellenkalkulationsdateien können die Namen der Variablen aus der ersten Zeile der Datei oder des angegebenen Bereichs eingelesen werden. Die Werte werden nach Bedarf umgewandelt, um gültige Variablennamen zu erstellen. Dabei werden Leerzeichen in Unterstriche umgewandelt.

Arbeitsblatt. Dateien von Excel 95 oder Nachfolgeversionen können mehrere Arbeitsblätter enthalten. In der Standardeinstellung liest der Daten-Editor das erste Arbeitsblatt. Wenn Sie ein anderes Arbeitsblatt einlesen möchten, wählen Sie es aus der Dropdown-Liste aus.

Bereich. Bei Datendateien aus Tabellenkalkulationen ist es außerdem möglich, nur einen bestimmten Zellenbereich einzulesen. Verwenden Sie beim Festlegen des Zellenbereichs dieselbe Methode wie im Tabellenkalkulationsprogramm.

Einlesen von Dateien aus Excel 95 oder nachfolgenden Versionen

Beim Einlesen von Dateien aus Excel 95 oder nachfolgenden Versionen gelten die folgenden Regeln:

Datentypen und Breiten. Jede Spalte stellt eine Variable dar. Der Datentyp und die Breite jeder Variablen werden durch den Datentyp und die Breite in der Excel-Datei bestimmt. Wenn eine Spalte mehr als einen Datentyp enthält (beispielsweise Datumsangaben und numerische Daten), wird der Datentyp auf "String" gesetzt und alle Werte werden als gültige String-Werte eingelesen.

Leere Zellen. Bei numerischen Variablen werden die leeren Zellen in systemdefinierte fehlende Werte konvertiert. Diese werden durch einen Punkt dargestellt. Bei String-Variablen stellen leere Zellen gültige Werte dar. Leere Zellen werden daher als gültige String-Variablen behandelt.

Variablennamen. Wenn aus der ersten Zeile der Excel-Datei (oder der ersten Zeile des angegebenen Bereichs) Variablennamen eingelesen werden, werden Werte, die nicht den Regeln für Variablennamen entsprechen, in gültige Variablennamen umgewandelt, und die ursprünglichen Spaltenüberschriften werden als Variablenlabels gespeichert. Falls keine Variablennamen aus der Excel-Datei eingelesen werden, erhalten die Variablen Standardnamen.

Einlesen von älteren Excel-Dateien und anderen Tabellenkalkulationsdateien

Beim Einlesen von Dateien aus älteren Excel-Versionen (vor Excel 95) und anderen Tabellenkalkulationsprogrammen gelten die folgenden Regeln:

Datentypen und Breiten. Die Datentypen und Breiten der Variablen werden durch die Datentypen und Spaltenbreiten der ersten Zellen mit Daten in den Spalten festgelegt. Werte anderer Typen werden in systemdefiniert fehlende Werte konvertiert. Wenn die erste Zelle mit Daten in einer Spalte leer ist, wird der globale Standarddatentyp des Tabellenkalkulationsblatts verwendet. In der Regel handelt es sich hierbei um einen numerischen Datentyp.

Leere Zellen. Bei numerischen Variablen werden die leeren Zellen in systemdefinierte fehlende Werte konvertiert. Diese werden durch einen Punkt dargestellt. Bei String-Variablen stellen leere Zellen gültige Werte dar. Leere Zellen werden daher als gültige String-Variablen behandelt.

Variablennamen. Wenn Sie die Namen der Variablen nicht aus dem Tabellenkalkulationsblatt einlesen, verwendet SPSS bei Excel- und Lotus-Dateien Buchstaben für die Bezeichnung der Spalten, also *A*, *B*, *C* usw., als Variablennamen. Bei SYLK-Dateien und Excel-Dateien, die im Anzeigeformat "R1C1" gespeichert wurden, verwendet SPSS den Buchstaben *C* und die Spaltennummer, also *C1*, *C2*, *C3* usw., als Variablennamen.

Einlesen von dBASE-Dateien

Datenbankdateien sind bezüglich der Logik ähnlich wie Datendateien im SPSS-Format aufgebaut. Bei dBASE-Dateien gelten die folgenden allgemeinen Regeln:

- Feldnamen werden in gültige Variablennamen umgewandelt.
- Falls in den Feldnamen in dBASE Doppelpunkte verwendet werden, werden diese in Unterstriche umgewandelt.
- Datensätze, die zwar zum Löschen markiert, aber noch nicht bereinigt wurden, werden berücksichtigt. SPSS erstellt eine neue String-Variable, *D_R*, und weist dieser bei Fällen, die zum Löschen markiert wurden, ein Sternchen zu.

Einlesen von Stata-Dateien

Bei Stata-Datendateien gelten die folgenden allgemeinen Regeln:

- **Variablennamen.** Stata-Variablennamen werden unter der Berücksichtigung der Groß- und Kleinschreibung in SPSS-Variablennamen umgewandelt. Stata-Variablennamen, die bis auf die Groß- und Kleinschreibung übereinstimmen, werden in gültige Variablennamen umgewandelt, indem ein Unterstrich und fortlaufende Buchstaben angehängt werden (*_A*, *_B*, *_C*, ..., *_Z*, *_AA*, *_AB*, ..., usw.).
- **Variablenlabels.** Stata-Variablenlabels werden in SPSS-Variablenlabels umgewandelt.
- **Wertelabels.** Mit Ausnahme der Stata-Wertelabels, die "erweiterten" fehlenden Werten zugewiesen sind, werden Stata-Wertelabels in SPSS-Wertelabels umgewandelt.
- **Fehlende Werte.** "Erweiterte" fehlende Werte aus Stata werden in systemdefiniert fehlende Werte umgewandelt.
- **Umwandlung von Datumsangaben.** Stata-Werte im Datumsformat werden in Werte mit dem SPSS-Format `DATE` (t-m-j) umgewandelt. Stata-Datumswerte für Zeitreihen (Wochen, Monate, Quartale usw.) werden in das einfache numerische Format (F) umgewandelt. Dabei bleibt der ursprüngliche, ganzzahlige Wert erhalten. Dies ist die Anzahl an Wochen, Monaten, Quartalen usw. seit dem Beginn des Jahres 1960.

Einlesen von Datenbankdateien

SPSS kann Daten aus Datenbankdateien in beliebigen Formaten einlesen, wenn Sie über einen entsprechenden Datenbanktreiber verfügen. Im Modus für lokale Analysen müssen die erforderlichen Treiber auf dem lokalen Computer installiert sein. Im Modus für verteilte Analysen (verfügbar mit SPSS Server) müssen die Treiber auf dem Remote-Server installiert sein. Für weitere Informationen siehe Modus für verteilte Analysen in Kapitel 4 auf S. 72.

So lesen Sie Datenbankdateien ein:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Datenbank öffnen
  Neue Abfrage...

- Wählen Sie die Datenquelle aus.
- Falls erforderlich (abhängig von der Datenquelle), wählen Sie die Datenbankdatei aus und/oder geben Sie einen Anmeldenamen, ein Passwort und andere Informationen ein.
- Wählen Sie die Tabelle(n) und Felder aus. Bei OLE DB-Datenquellen (nur unter Windows-Betriebssystemen verfügbar) können Sie nur eine Tabelle auswählen.
- Legen Sie gegebenenfalls Relationen zwischen den Tabellen fest.
- Die folgenden Optionen sind verfügbar:
  - Auswahlkriterien für die Daten festlegen,
  - Eine Aufforderung für benutzerdefinierte Eingaben hinzufügen, um eine Parameterabfrage zu erstellen,
  - Speichern Sie die erstellte Abfrage, bevor Sie sie ausführen.

So können Sie gespeicherte Datenbankabfragen bearbeiten:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Datenbank öffnen
  Abfrage bearbeiten...

- Wählen Sie die Abfragedatei (**.spq*) aus, die Sie bearbeiten möchten.
- Zum Erstellen einer neuen Abfrage folgen Sie den Anweisungen.

So lesen Sie Datenbankdateien mit gespeicherten Abfragen ein:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Datenbank öffnen
  Abfrage ausführen...

- Wählen Sie die auszuführende Abfragedatei (**.spq*) aus.
- Falls erforderlich (abhängig von der Datenbankdatei), geben Sie einen Anmeldenamen und ein Passwort ein.
- Wenn für die Abfrage eine Eingabeaufforderung definiert wurde, müssen Sie ggf. weitere Informationen eingeben (beispielsweise das Quartal, für das Sie die Verkaufszahlen abrufen möchten).

Auswählen einer Datenquelle

Wählen Sie im ersten Bildschirm des Datenbank-Assistenten den Typ der einzulesenden Datenquelle aus.

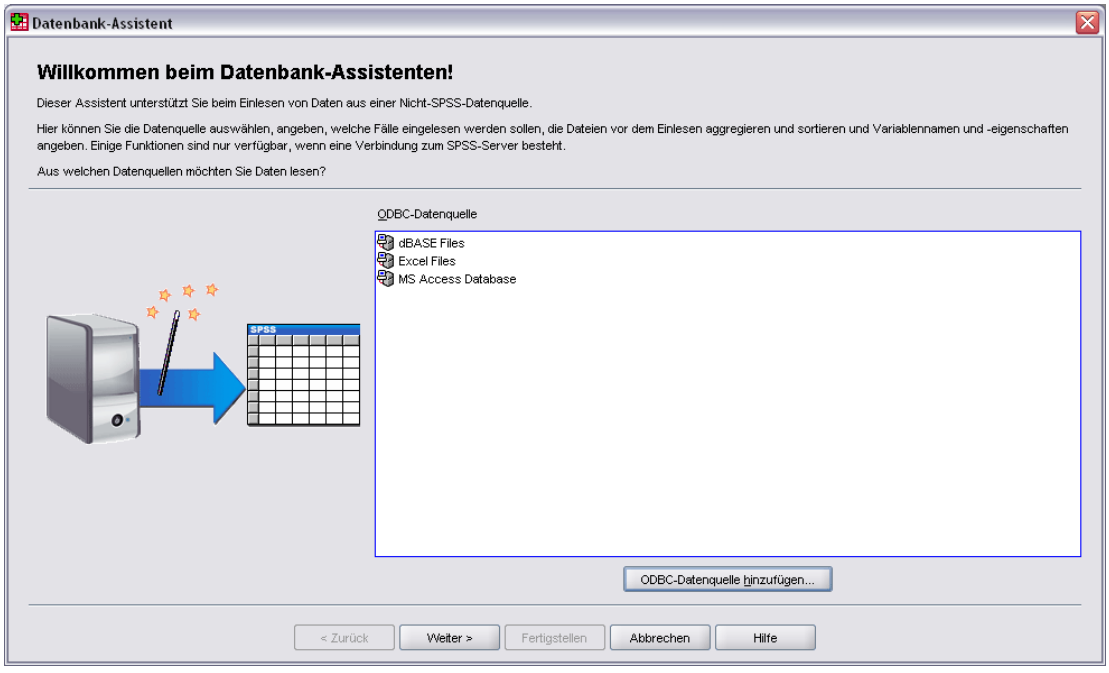
ODBC-Datenquellen

Wenn Sie noch keine ODBC-Datenquelle konfiguriert haben oder eine neue Datenquelle hinzufügen möchten, klicken Sie auf ODBC-Datenquelle hinzufügen.

- Bei Linux-Betriebssystemen ist diese Schaltfläche nicht verfügbar. ODBC-Datenquellen werden in *odbc.ini* angegeben und für die *ODBCINI*-Umgebungsvariablen muss der Speicherort der betreffenden Datei festgelegt sein. Weitere Informationen finden Sie in der Hilfe zu Ihren Datenbanktreibern.
- Im Modus für verteilte Analysen (verfügbar mit SPSS Server) steht diese Schaltfläche nicht zur Verfügung. Wenn Sie Datenquellen im Modus für verteilte Analysen hinzufügen möchten, wenden Sie sich an Ihren Systemadministrator.

Eine ODBC-Datenquelle besteht aus zwei wichtigen Informationen: dem Treiber, der zum Zugreifen auf die Daten verwendet wird, und dem Speicherort der Datenbank, auf die Sie zugreifen möchten. Wenn Sie Datenquellen definieren möchten, muss der entsprechende Treiber installiert sein. Für den Modus für lokale Analysen können Sie Treiber für eine Vielzahl von Datenbankformaten von der SPSS-Installations-CD-ROM installieren.

Abbildung 3-1
Datenbank-Assistent



OLE DB-Datenquellen

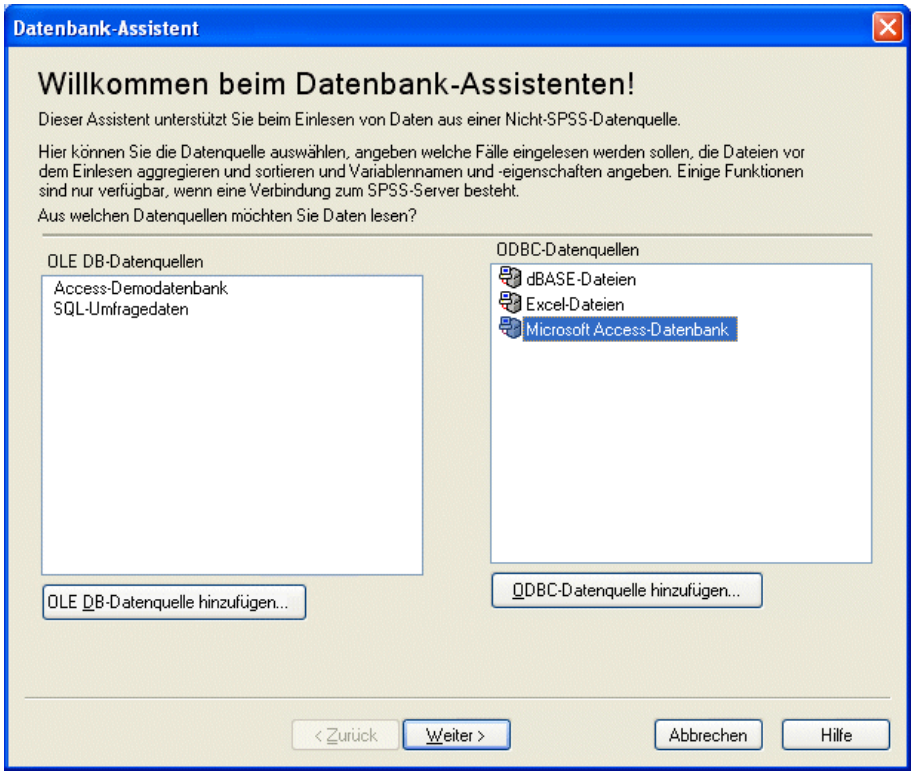
Wenn Sie auf OLE DB-Datenquellen zugreifen möchten (nur unter Microsoft Windows-Betriebssytemen verfügbar), müssen folgende Komponenten installiert sein:

- .NET Framework
- Dimensions-Datenmodell und OLE DB Access

Die Versionen dieser Komponenten, die mit dieser Version kompatibel sind, können von der Installations-CD installiert werden und sind über das AutoPlay-Menü verfügbar.

- Verbindungen zwischen Tabellen sind für OLE-DB-Datenquellen nicht verfügbar. Es kann jeweils nur eine Tabelle gelesen werden.
- OLE DB-Datenquellen können nur im Modus für lokale Analysen hinzugefügt werden. Falls Sie OLE DB-Datenquellen auf einem Windows-Server im Modus für verteilte Analysen hinzufügen möchten, wenden Sie sich an Ihren Systemadministrator.
- Im Modus für verteilte Analysen (verfügbar mit SPSS Server) sind OLE DB-Datenquellen nur auf Windows-Servern verfügbar. Außerdem müssen sowohl .NET als auch das Dimensions-Datenmodell sowie OLE DB Access auf dem Server installiert sein.

Abbildung 3-2
Datenbank-Assistent mit Zugriff auf OLE DB-Datenquellen

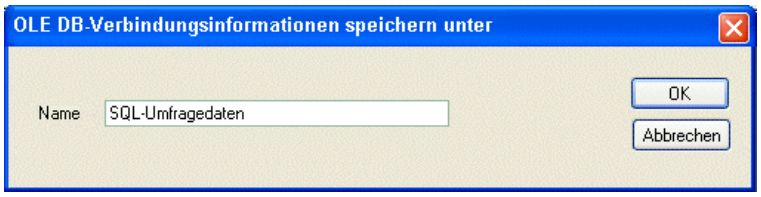


So fügen Sie eine OLE DB-Datenquelle hinzu:

► Klicken Sie auf OLE DB-Datenquelle hinzufügen.

- Klicken Sie unter “Eigenschaften der Datenverknüpfung” auf die Registerkarte Provider und wählen Sie den OLE DB-Provider aus.
- Klicken Sie auf Weiter oder klicken Sie auf die Registerkarte Verbindung.
- Wählen Sie die Datenbank aus, indem Sie das Verzeichnis und den Datenbanknamen eingeben, oder indem Sie auf die Schaltfläche klicken, um das Verzeichnis nach einer Datenbank zu durchsuchen. (Möglicherweise ist ein Benutzername und ein Passwort erforderlich.)
- Klicken Sie auf OK, nachdem Sie die erforderlichen Informationen eingegeben haben. (Durch Klicken auf die Schaltfläche Verbindung prüfen können Sie sicherstellen, dass die angegebene Datenbank verfügbar ist.)
- Geben Sie einen Namen für die Informationen zur Datenbankverbindung ein. (Dieser Name wird in der Liste der verfügbaren OLE DB-Datenquellen angezeigt.)

Abbildung 3-3
Dialogfeld “OLE DB-Verbindungsinformationen speichern unter”



- Klicken Sie auf OK.

Hiermit gelangen Sie zurück zum ersten Bildschirm des Datenbank-Assistenten, auf dem Sie den gespeicherten Namen aus der Liste der OLE DB-Datenquellen auswählen und mit den weiteren Schritten des Assistenten fortfahren können.

Löschen von OLE DB-Datenquellen

Um Datenquellennamen aus der Liste der OLE DB-Datenquellen zu löschen, müssen Sie die UDL-Datei mit dem Namen der Datenquelle in folgendem Verzeichnis löschen:

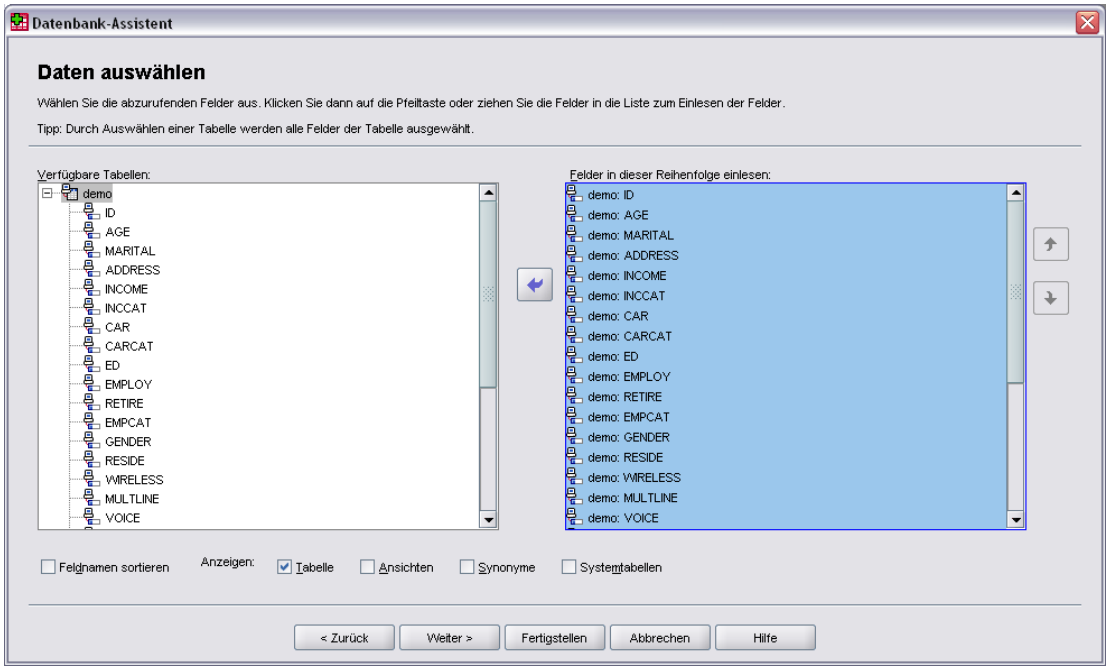
[Laufwerk:\Dokumente und Einstellungen\[Benutzername]\Lokale Einstellungen\Anwendungsdaten\SPSS\UDL

Auswählen von Datenfeldern

Mit dem Schritt “Daten auswählen” wird gesteuert, welche Tabellen und Felder eingelesen werden sollen. Datenbankfelder (Spalten) werden als Variablen in SPSS eingelesen.

Wenn in einer Tabelle eine beliebige Anzahl von Feldern ausgewählt wurde, werden im nächsten Fenster des Datenbank-Assistenten alle Felder dieser Tabelle angezeigt. Es werden jedoch nur die in diesem Schritt ausgewählten Felder als Variablen importiert. Auf diese Weise können Sie Verbindungen zwischen Tabellen erstellen und Kriterien festlegen, indem Sie die Felder verwenden, die nicht importiert werden.

Abbildung 3-4
Datenbank-Assistent, Auswählen von Daten



Anzeigen von Feldnamen. Klicken Sie auf das Pluszeichen (+) links neben dem Namen einer Tabelle, wenn Sie die Felder in dieser Tabelle einblenden möchten. Klicken Sie auf das Minuszeichen (–) links neben dem Namen einer Tabelle, wenn Sie die Felder ausblenden möchten.

So fügen Sie ein Feld hinzu: Doppelklicken Sie in der Liste “Verfügbare Tabellen” auf das entsprechende Feld oder ziehen Sie dieses in das Listenfeld “Felder in dieser Reihenfolge einlesen”. Sie können die Reihenfolge der Felder ändern, indem Sie diese in der Liste der Felder ziehen und ablegen.

So entfernen Sie ein Feld: Doppelklicken Sie in der Liste “Felder in dieser Reihenfolge einlesen” auf das entsprechende Feld oder ziehen Sie dieses in das Listenfeld “Verfügbare Tabellen”.

Feldnamen sortieren. Wenn dieses Kontrollkästchen aktiviert ist, werden die verfügbaren Felder im Datenbank-Assistenten in alphabetischer Reihenfolge angezeigt.

Standardmäßig werden in der Liste der verfügbaren Tabellen nur Standarddatenbanktabellen angezeigt. Sie können festlegen, welche Art von Elementen in der Liste angezeigt werden sollen:

- **Tabellen.** Standarddatenbanktabellen.
- **Ansichten.** Ansichten sind virtuelle oder dynamische “Tabellen”, die durch Abfragen definiert werden. Dazu können Verknüpfungen mehrerer Tabellen und/oder Felder gehören, die aus Berechnungen auf der Grundlage der Werte in anderen Feldern abgeleitet wurden.

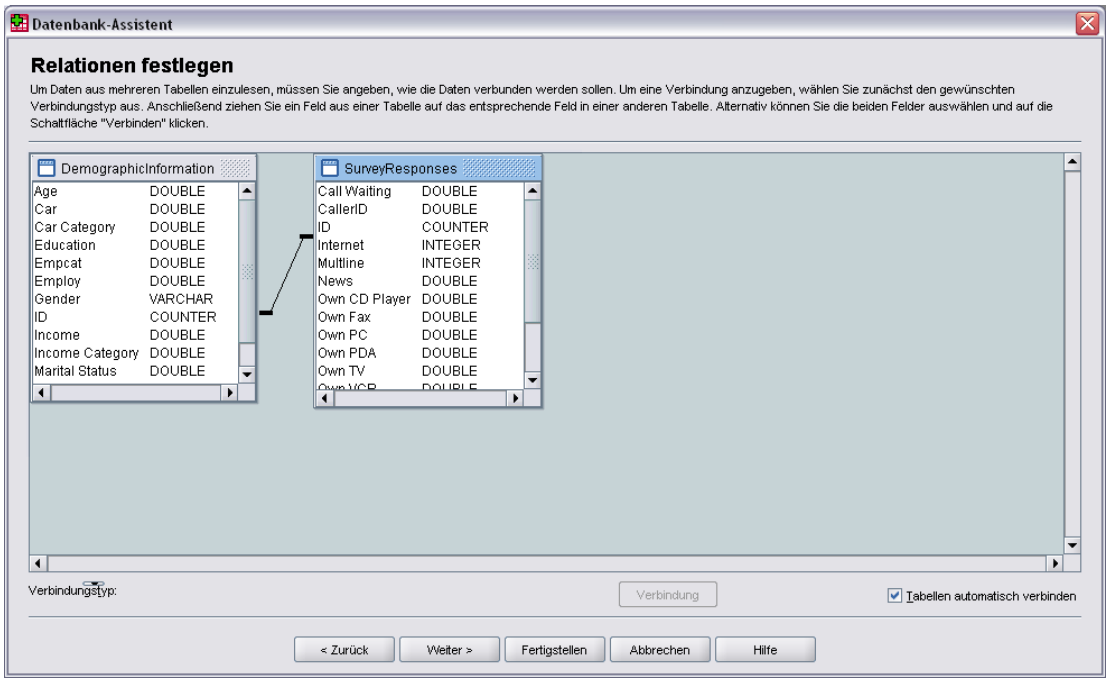
- **Synonyme.** Ein Synonym ist ein Alias für eine Tabelle oder eine Ansicht und wird normalerweise in einer Abfrage definiert.
- **Systemtabellen.** Systemtabellen definieren Datenbankeigenschaften. In einigen Fällen können Standarddatenbanktabellen als Systemtabellen klassifiziert sein und nur bei Auswahl dieser Option angezeigt werden. Der Zugriff auf eigentliche Systemtabellen ist häufig auf Datenbankadministratoren beschränkt.

Anmerkung: Bei OLE DB-Datenquellen (nur unter Windows-Betriebssystemen verfügbar) können Sie Felder nur aus einer einzigen Tabelle auswählen. Verknüpfungen zwischen mehreren Tabellen werden bei OLE-DB-Datenquellen nicht unterstützt.

Erstellen einer Beziehung zwischen Tabellen

Im Schritt “Relationen festlegen” können Sie für ODBC-Datenquellen die Relationen zwischen den Tabellen festlegen. Wenn Felder aus mehr als einer Tabelle ausgewählt sind, müssen Sie mindestens eine Verbindung festlegen.

Abbildung 3-5
Datenbank-Assistent, Festlegen von Relationen



Herstellen von Relationen. Zum Erstellen von Relationen ziehen Sie ein Feld aus einer beliebigen Tabelle auf das Feld, mit dem Sie dieses verbinden möchten. Im Datenbank-Assistenten wird dann eine **Verbindungslinie** zwischen den beiden Feldern angezeigt. Diese stellt die Beziehung dar. Die Datentypen der beiden Felder müssen übereinstimmen.

Tabellen automatisch verbinden. Hierbei wird versucht, Tabellen anhand von Primär-/Fremdschlüsseln oder übereinstimmenden Feldnamen und Datentypen automatisch zu verbinden.

Verbindungstyp. Wenn der von Ihnen eingesetzte Treiber äußere Verbindungen unterstützt, können Sie innere, linke äußere und rechte äußere Verbindungen festlegen.

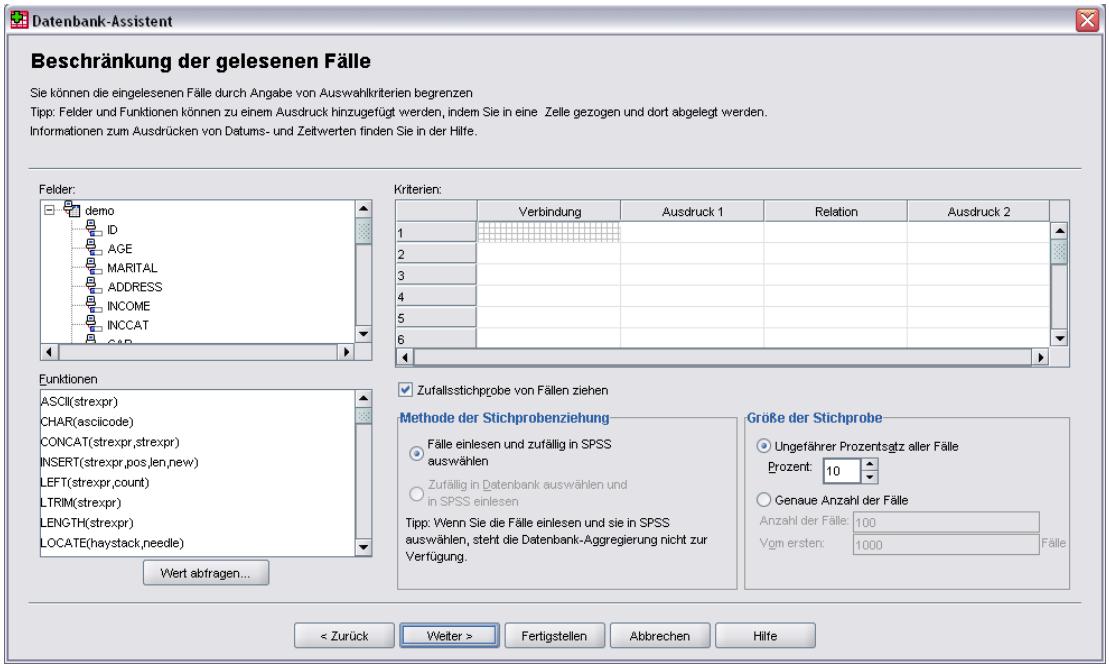
- **Innere Verknüpfungen.** Eine innere Verknüpfung enthält nur die Zeilen, bei denen die verbundenen Felder übereinstimmen. In diesem Beispiel werden alle Zeilen mit übereinstimmenden ID-Werten in beiden Tabellen berücksichtigt.
- **Äußere Verknüpfungen.** Mit einer inneren Verknüpfung lassen sich Zuordnungen in Tabellen herstellen, die in einer 1:1-Beziehung stehen. Mit einer äußeren Verknüpfung sind zusätzlich Zuordnungen von Tabellen in einer 1:n-Beziehung möglich. So können Sie beispielsweise eine Tabelle mit nur wenigen Datensätzen, die Datenwerte und entsprechende beschreibende Labels darstellen, einer Tabelle mit hunderten oder tausenden Datensätzen zuordnen, die Teilnehmer an einer Umfrage darstellen. Eine linke äußere Verknüpfung enthält alle Datensätze aus der Tabelle auf der linken Seite und nur die Datensätze aus der Tabelle auf der rechten Seite, bei denen die verknüpften Felder übereinstimmen. In einer rechten äußeren Verbindung werden alle Datensätze aus der Tabelle auf der rechten Seite und nur die Datensätze aus der Tabelle auf der linken Seite importiert, bei denen die verknüpften Felder übereinstimmen.

Beschränkung der gelesenen Fälle

Im Schritt “Beschränkung der gelesenen Fälle” können Sie Kriterien festlegen, mit denen Teilmengen von Fällen (Zeilen) ausgewählt werden. Im Allgemeinen werden zum Beschränken von Fällen die Kriterien in die Kriterientabelle eingegeben. Kriterien bestehen aus zwei Ausdrücken und einer zwischen diesen festgelegten Beziehung. Die Ausdrücke geben für jeden Fall die Werte *Wahr*, *Falsch* oder *Fehlend* zurück.

- Wenn als Ergebnis der Wert *Wahr* vorliegt, wird der Fall ausgewählt.
- Wenn als Ergebnis der Wert *Falsch* oder *Fehlend* vorliegt, wird der Fall nicht ausgewählt.
- Bei den meisten Kriterien wird mindestens einer der sechs Vergleichsoperatoren (<, >, <=, >=, =, <>) verwendet.
- Bedingte Ausdrücke können Feldnamen, Konstanten, arithmetische Operatoren, numerische und andere Funktionen sowie logische Variablen enthalten. Sie können Felder, die nicht importiert werden sollen, als Variablen verwenden.

Abbildung 3-6
Datenbank-Assistent, Beschränkung der gelesenen Fälle



Wenn Sie ein Kriterium erstellen möchten, benötigen Sie mindestens zwei Ausdrücke und eine zwischen den Ausdrücken festgelegte Beziehung, um diese Ausdrücke zu verbinden.

- Wenn Sie einen Ausdruck erstellen möchten, wählen Sie eine der folgenden Methoden aus:
  - Geben Sie in einer Ausdruckszelle Feldnamen, Konstanten, arithmetische Operatoren, numerische und andere Funktionen oder logische Variablen ein.
  - Doppelklicken Sie in der Liste “Felder” auf das Feld.
  - Ziehen Sie das Feld aus der Liste “Felder” auf eine Ausdruckszelle.
  - Wählen Sie aus dem Dropdown-Menü einer beliebigen aktiven Ausdruckszelle ein Feld aus.
- Um einen relationalen Operator (z. B. = oder >) auszuwählen, setzen Sie die Einfügemarke in die Beziehungszelle und geben Sie entweder den Operator manuell ein oder wählen Sie diesen im Dropdown-Menü aus.

Wenn die SQL `WHERE`-Klauseln mit Ausdrücken für die Fallauswahl enthält, müssen die Datums- und Zeitangaben in Ausdrücken auf spezielle Weise angegeben werden (einschließlich der in den Beispielen zu sehenden geschweiften Klammern):

- Für Datumsangaben gilt die folgende allgemeine Form `{d 'jjj-mm-tt'}`.
- Für Uhrzeitangaben gilt die folgende allgemeine Form `{t 'hh:mm:ss'}`.

- Für Datums- und Uhrzeitangaben (Zeitstempel) gilt die folgende allgemeine Form: `{ts 'jjjj-mm-tt hh:mm:ss'}`.
- Der gesamte Datums- und/oder Zeitwert muss in einfache Anführungsstriche eingeschlossen sein. Jahre müssen in vierstelliger Form angegeben werden und Datums- und Uhrzeitangaben müssen für jeden Bereich des Werts zwei Ziffern enthalten. Der erste Januar 2005, 1:05 Uhr würde also wie folgt angegeben:

  ```
  {ts '2005-01-01 01:05:00'}
  ```

Funktionen. SPSS stellt eine Reihe von arithmetischen und logischen SQL-Funktionen sowie SQL-Funktionen für Zeichenfolgen, Datumsangaben und Zeitangaben zur Verfügung. Sie können diese Funktionen aus der Liste auswählen und in den Ausdruck ziehen oder beliebige gültige SQL-Funktionen eingeben. Informationen zu den gültigen SQL-Funktionen finden Sie in der Dokumentation Ihrer Datenbank. Eine Liste der Standardfunktionen finden Sie unter:

http://msdn2.microsoft.com/en-us/library/ms711813.aspx

Zufallsstichproben verwenden. Mit dieser Option wird aus der Datenquelle eine Zufallsstichprobe von Fällen ausgewählt. Bei großen Datenquellen soll die Anzahl der Fälle möglicherweise auf eine kleine, repräsentative Auswahl begrenzt werden, womit die Laufzeit von Prozeduren beträchtlich verringert werden kann. Integrierte Zufallsstichproben sind, falls für die Datenquelle verfügbar, schneller als SPSS -Zufallsstichproben, da bei SPSS -Zufallsstichproben noch die gesamte Datenquelle gelesen werden muss, um eine Zufallsstichprobe zu extrahieren.

- **Ungefähr.** Erstellt eine Zufallsstichprobe, die ungefähr den angegebenen Prozentsatz aller Fälle enthält. Da diese Routine für jeden Fall eine unabhängige Pseudo-Zufallsentscheidung trifft, entspricht der Prozentsatz der tatsächlich ausgewählten Fälle dem angegebenen Prozentwert nur ungefähr. Je mehr Fälle sich in der Datendatei befinden, desto eher entspricht der Prozentsatz ausgewählter Fälle dem angegebenen Prozentsatz.
- **Exakt.** Wählt eine Zufallsstichprobe mit der angegebenen Anzahl von Fällen aus der festgelegten Gesamtanzahl der Fälle aus. Wenn die angegebene Gesamtanzahl der Fälle größer als die Anzahl der Fälle in der Datendatei ist, enthält die Zufallsstichprobe proportional weniger Fälle als angefordert wurden.

Anmerkung: Bei Zufallsstichproben steht die Aggregation (verfügbar im Modus für verteilte Analysen mit SPSS Server) nicht zur Verfügung.

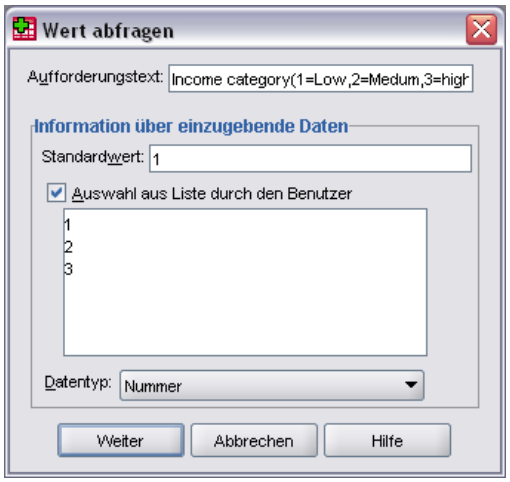
Wert abfragen. Zum Erstellen einer **Parameterabfrage** können Sie in die Abfrage eine Eingabeaufforderung integrieren. Bei Ausführen der Abfrage werden die Benutzer dann aufgefordert, anhand der dieser Angaben Informationen einzugeben. Eingabeaufforderungen können in Situationen nützlich sein, in denen verschiedene Ansichten derselben Daten benötigt werden. Sie möchten beispielsweise die Verkaufszahlen für verschiedene Rechnungsjahre unter Verwendung derselben Abfrage einsehen.

▶ Setzen Sie die Einfügemarke in eine beliebige Ausdruckszelle und klicken Sie zum Erstellen einer Eingabeaufforderung auf Wert abfragen.

Erstellen einer Parameterabfrage

Im Schritt “Wert abfragen” können Sie ein Dialogfeld erstellen, in dem die Benutzer bei jedem neuen Ausführen einer Abfrage aufgefordert werden, die entsprechenden Informationen einzugeben. Diese Funktion ist nützlich, wenn dieselbe Datenquelle anhand von verschiedenen Kriterien abgefragt werden soll.

Abbildung 3-7
Wert abfragen



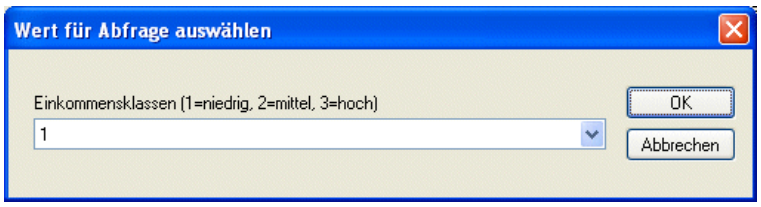
Geben Sie zum Erstellen einer Eingabeaufforderung den Aufforderungstext und einen Standardwert ein. Der Aufforderungstext wird immer dann angezeigt, wenn ein Benutzer die Abfrage ausführt. Der Text sollte die Art der einzugebenden Information beschreiben. Wenn der Benutzer nicht aus einer Liste auswählt, sollte der Text einen Hinweis darauf geben, wie die Eingabe formatiert werden soll. Ein Beispiel lautet folgendermaßen: Geben Sie ein Quartal ein (Q1, Q2, Q3, ...).

Auswahl aus Liste durch den Benutzer. Wenn dieses Kontrollkästchen aktiviert ist, wird der Benutzer auf die Auswahl der von Ihnen bereitgestellten Werte beschränkt. Stellen Sie sicher, dass die Werte durch Zeilenumbrüche getrennt sind.

Datentyp. Wählen Sie hier den zu verwendenden Datentyp aus (Numerisch, String oder Datum).

Das abschließende Ergebnis sieht folgendermaßen aus:

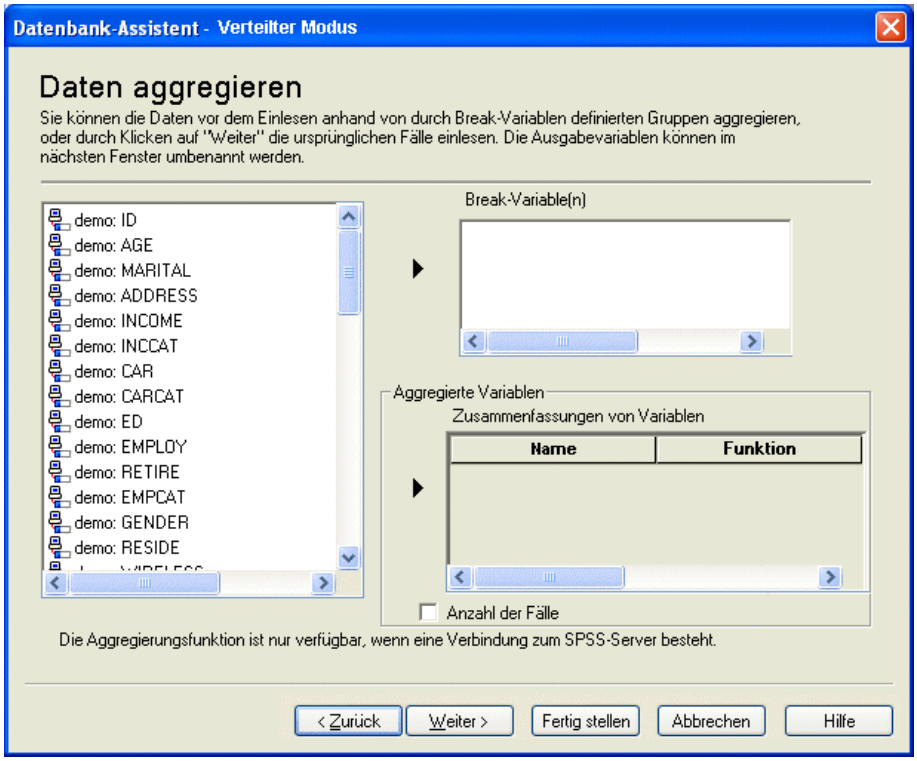
Abbildung 3-8
Benutzerdefinierte Eingabeaufforderung



Aggregieren von Daten

Wenn Sie sich im Modus für verteilte Analysen befinden und eine Verbindung zu einem Remote-Server besteht (mit SPSS Server verfügbar), können Sie die Daten vor dem Einlesen in SPSS aggregieren.

Abbildung 3-9
Datenbank-Assistent, Aggregieren von Daten



Die Daten können auch nach dem Einlesen in SPSS aggregiert werden; bei umfangreichen Datenquellen kann die vorherige Aggregierung jedoch beträchtlich Zeit sparen.

- Um aggregierte Daten zu erstellen, wählen Sie eine oder mehrere Break-Variablen aus, die definieren, wie die Fälle gruppiert werden.
- Wählen Sie mindestens eine aggregierte Variable aus.
- Wählen Sie für jede Aggregierungsvariable eine Aggregierungsfunktion aus.
- Sie können bei Bedarf auch eine Variable anlegen, welche die Anzahl der Fälle in jeder Break-Gruppe enthält.

  Anmerkung: Bei SPSS -Zufallsstichproben ist die Aggregation nicht verfügbar.

Definieren von Variablen

Variablennamen und -labels. In SPSS wird der vollständige Name des Datenbankfelds (der Spalte) als Variablenlabel verwendet. Wenn Sie keine Änderungen an den Variablennamen vornehmen, weist der Datenbank-Assistent jeder Spalte der Datenbank selbständig einen Variablennamen zu. Bei der Vergabe von Variablennamen werden die beiden folgenden Verfahren eingesetzt:

- Wenn der Name des Datenbankfelds einen gültigen und eindeutigen Variablennamen für SPSS ergibt, wird dieser Name als Variablenname verwendet.
- Wenn der Name des Datenbankfelds keinen gültigen und eindeutigen Variablennamen ergibt, wird automatisch ein neuer, eindeutiger Name erstellt.

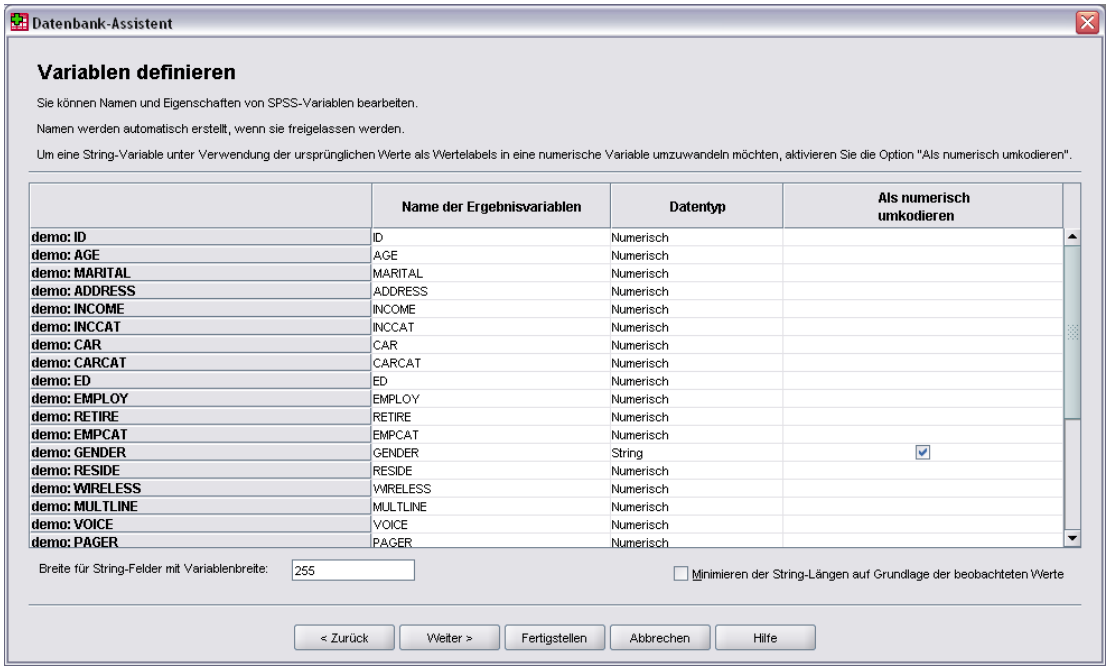
Klicken Sie auf eine beliebige Zelle, um den Variablennamen zu bearbeiten.

Umwandeln von Strings in numerische Werte. Wählen Sie das Feld Als numerisch umkodieren bei einer String-Variablen aus, wenn diese automatisch in eine numerische Variable umgewandelt werden soll. String-Werte werden anhand der alphabetischen Reihenfolge der ursprünglichen Werte in fortlaufende, ganzzahlige Werte umgewandelt. Die ursprünglichen Werte werden als Wertelabels für die neuen Variablen beibehalten.

Breite für String-Felder mit Variablenbreite. Mit dieser Option wird die Breite der String-Werte mit Variablenbreite gesteuert. Standardmäßig beträgt die Breite 255 Byte und nur die ersten 255 Byte (in der Regel 255 Zeichen bei Single-Byte-Sprachen) werden gelesen. Die Breite kann bis zu 32.767 Byte umfassen. In der Regel sollen String-Werte zwar nicht gekürzt werden, aber auch übermäßig große Werte sollten vermieden werden, weil dies zu Leistungseinbußen bei der Verarbeitung führt.

String-Längen anhand beobachteter Werte minimieren. Setzt die Länge der einzelnen String-Variablen auf den längsten beobachteten Wert.

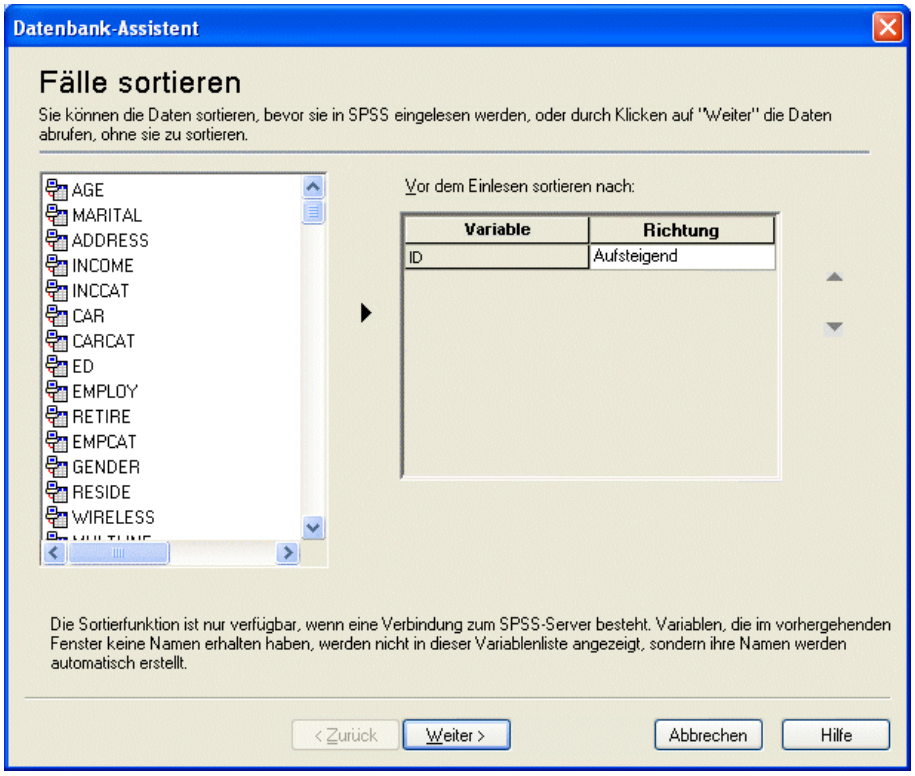
Abbildung 3-10
Datenbank-Assistent, Definieren von Variablen



Sortieren von Fällen

Wenn Sie sich im Modus für verteilte Analysen befinden und eine Verbindung zu einem Remote-Server besteht (mit SPSS Server verfügbar), können Sie die Daten vor dem Einlesen in SPSS sortieren.

Abbildung 3-11
Datenbank-Assistent, Sortieren von Fällen



Die Daten können auch nach dem Einlesen in SPSS sortiert werden; bei umfangreichen Datenquellen kann die vorherige Sortierung jedoch beträchtlich Zeit sparen.

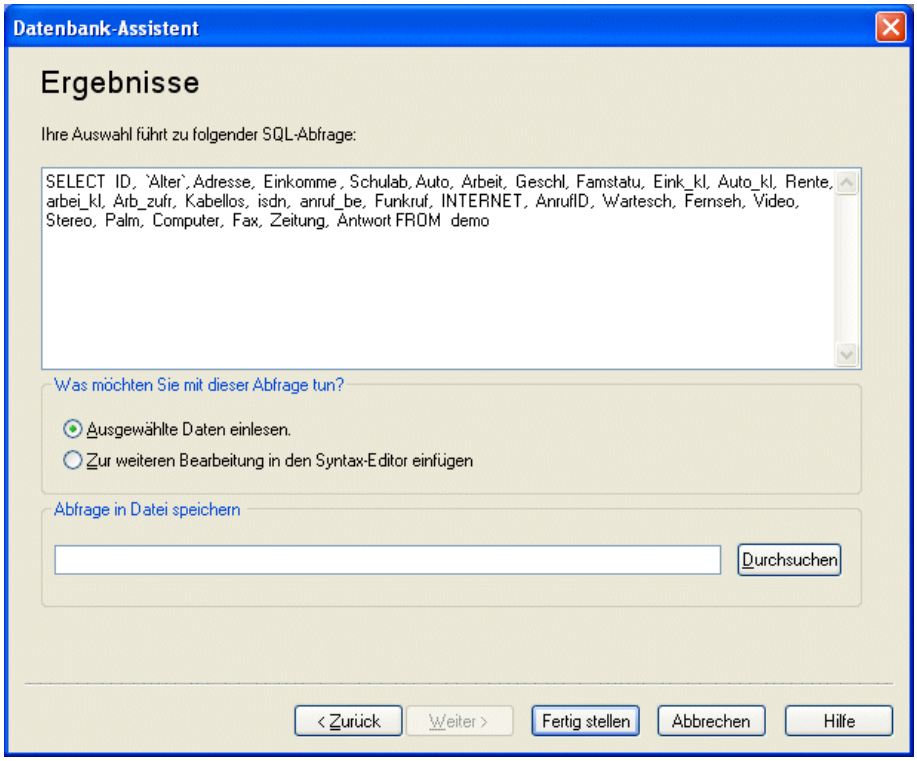
Ergebnisse

Im Schritt “Ergebnisse” wird die SQL-SELECT-Anweisung der Abfrage angezeigt.

- Sie können die SQL-SELECT-Anweisung vor dem Ausführen der Abfrage bearbeiten. Wenn Sie jedoch auf die Schaltfläche Zurück klicken, um Änderungen in vorangegangenen Schritten vorzunehmen, gehen die Änderungen an der SELECT-Anweisung verloren.
- Verwenden Sie den Abschnitt Abfrage in Datei speichern, um die Abfrage für eine zukünftige Verwendung zu speichern.
- Um die vollständige Syntax für `GET DATA` in ein Syntax-Fenster einzufügen, wählen Sie Zur weiteren Bearbeitung in den Syntax-Editor einfügen aus. Durch einfaches Kopieren und Einfügen der SELECT-Anweisung aus dem Ergebnisfeld wird nicht die erforderliche Befehlssyntax eingefügt.

Anmerkung: Die eingefügte Syntax enthält in jeder SQL-Zeile, die vom Assistenten erzeugt wurde, ein Leerzeichen vor dem schließenden Anführungszeichen. Diese Leerzeichen sind nicht überflüssig. Bei der Verarbeitung des Befehls werden alle Zeilen in der SQL-Anweisung zusammengeführt. Ohne das Leerzeichen würde das letzte Zeichen in einer Zeile übergangslos am ersten Zeichen in der jeweils nachfolgenden Zeile “kleben”.

Abbildung 3-12
Datenbank-Assistent, Ergebnisfeld



Text-Assistent

Der Text-Assistent kann Textdatendateien in einer Vielzahl von verschiedenen Formaten einlesen:

- Tabulator-getrennte Dateien
- Leerzeichen-getrennte Dateien
- Komma-getrennte Dateien
- Dateien mit Spalten fester Breite

Sie können bei Dateien mit Trennzeichen auch andere Zeichen als Trennzeichen zwischen Werten festlegen, und Sie können mehrere Trennzeichen bestimmen.

So lesen Sie Textdatendateien ein

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

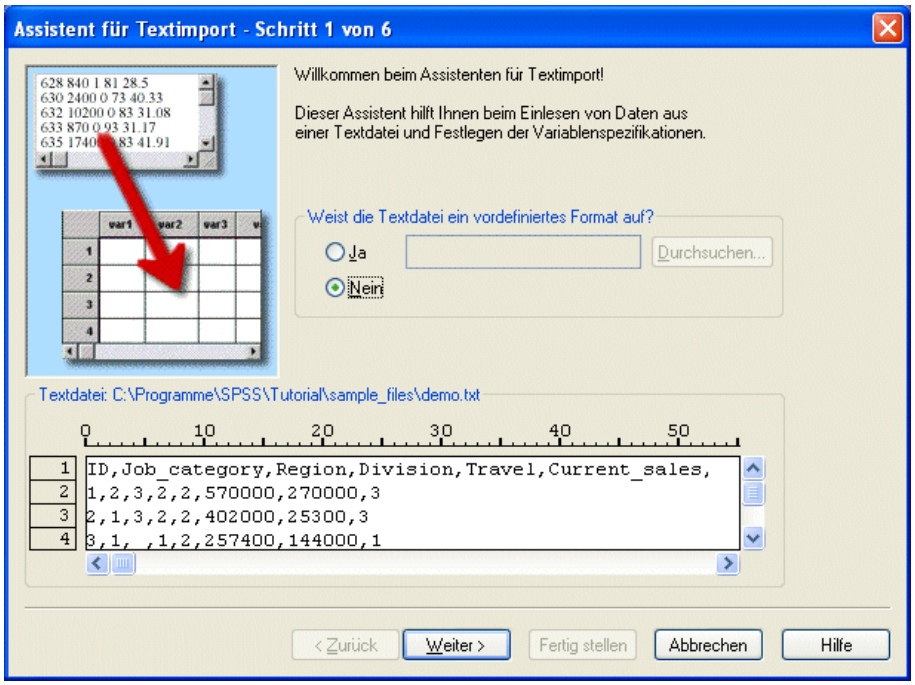
Datei
  Textdaten einlesen...

▶ Wählen Sie im Dialogfeld “Dateien öffnen” die Textdatei aus.

▶ Bestimmen Sie, wie die Datendatei eingelesen werden soll. Gehen Sie dazu anhand der Schritte im Text-Assistenten vor.

Text-Assistent: Schritt 1

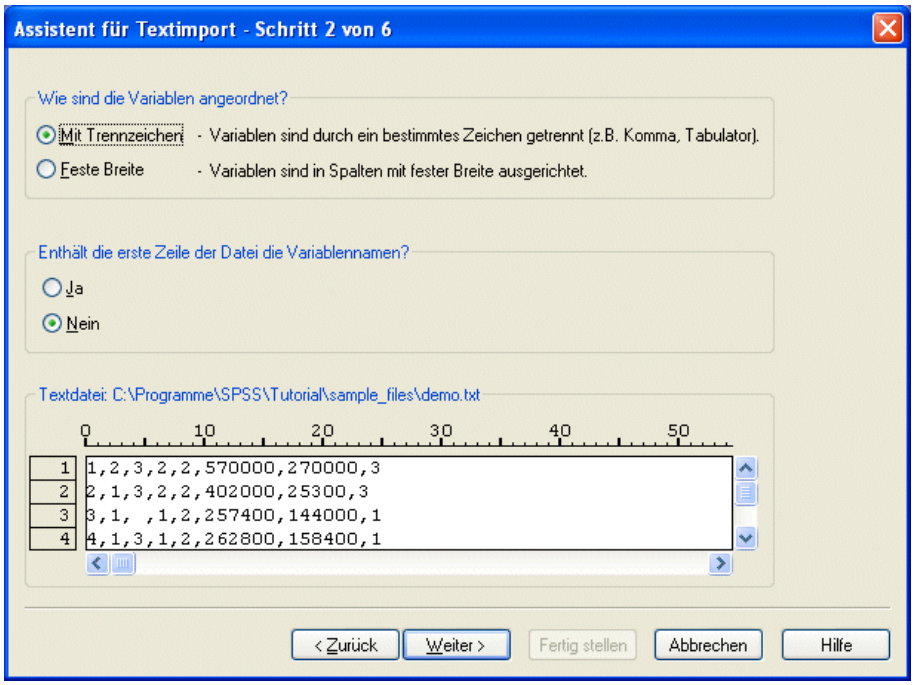
Abbildung 3-13
Text-Assistent: Schritt 1



Die Textdatei wird im Vorschaufenster angezeigt. Sie können ein vordefiniertes (also zuvor im Text-Assistenten gespeichertes) Format anwenden oder anhand der Schritte im Text-Assistenten vorgehen, um festzulegen, wie die Daten eingelesen werden sollen.

Text-Assistent: Schritt 2

Abbildung 3-14
Text-Assistent: Schritt 2



In diesem Schritt erhalten Sie Informationen über die Variablen. Eine Variable ist vergleichbar mit einem Feld in einer Datenbank. So stellt beispielsweise jede Frage auf einem Fragebogen eine Variable dar.

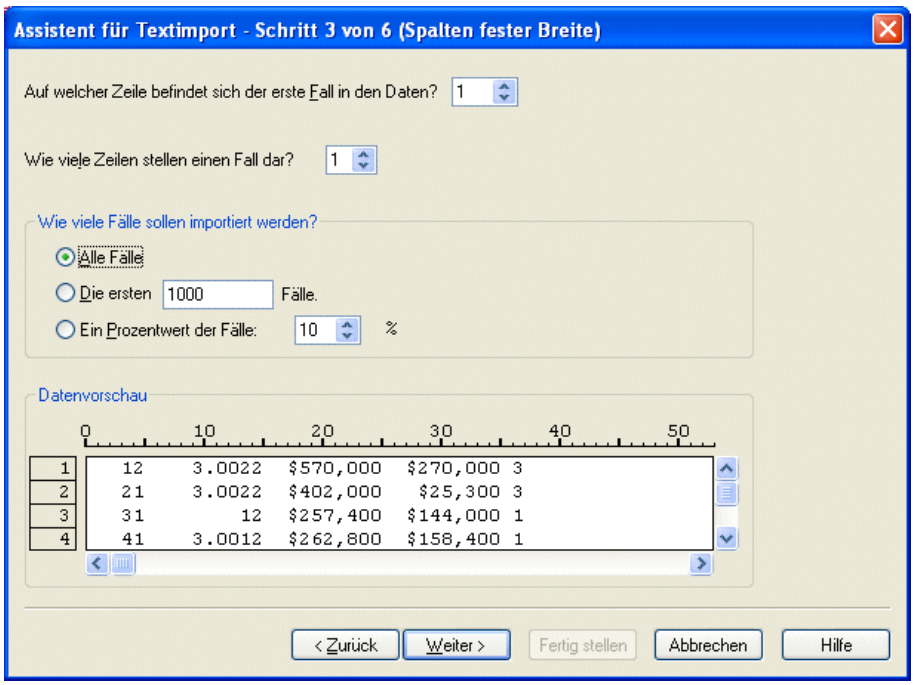
Wie sind die Variablen angeordnet? Damit der Text-Assistent die Daten ordnungsgemäß einlesen kann, müssen Sie dem Text-Assistenten mitteilen, woran das Ende des Datenwerts einer Variablen bzw. der Anfang des Datenwerts der nachfolgenden Variablen erkannt werden kann. Die Methode, mit der eine Variable von der nächsten unterschieden wird, ist abhängig von der Anordnung der Variablen.

- **Mit Trennzeichen.** Leerzeichen, Kommata, Tabulatoren oder andere Zeichen werden zum Trennen von Variablen verwendet. Die Variablen werden für jeden Fall in derselben Reihenfolge aufgezeichnet, befinden sich aber nicht notwendigerweise an derselben Spaltenposition.
- **Feste Breite.** Jede Variable wird für jeden Fall in der Datendatei an der gleichen Spaltenposition in derselben Zeile im Datensatz aufgezeichnet. Zwischen den Variablen sind keine Trennzeichen erforderlich. Bei vielen durch Computerprogramme erzeugten Textdatendateien kann es den Anschein haben, dass die Datenwerte ohne Trennzeichen ineinanderlaufen. Durch die Spaltenposition ist jedoch festgelegt, welche Variable eingelesen wird.

Enthält die erste Zeile der Datei die Variablennamen? Wenn die erste Zeile der Datendatei aussagekräftige Beschreibungen für jede Variable enthält, können Sie diese Beschreibungen als Variablennamen einsetzen. Werte, die nicht den Regeln für Variablennamen entsprechen, werden in gültige Variablennamen umgewandelt.

Text-Assistent: Schritt 3 (durch Trennzeichen getrennte Dateien)

Abbildung 3-15
Text-Assistent: Schritt 3 (für durch Trennzeichen getrennte Dateien)



In diesem Schritt erhalten Sie Informationen über die Fälle. Ein Fall ist vergleichbar mit einem Datensatz in einer Datenbank. So ist zum Beispiel jede Person, die einen Fragebogen ausfüllt, ein Fall.

Auf welcher Zeile befindet sich der erste Fall in den Daten? Gibt die erste Zeile der Datendatei an, die Datenwerte enthält. Wenn die erste(n) Zeile(n) der Datendatei aussagekräftige Beschreibungen oder anderen Text und somit keine Datenwerte enthalten, ist dies *nicht* Zeile 1.

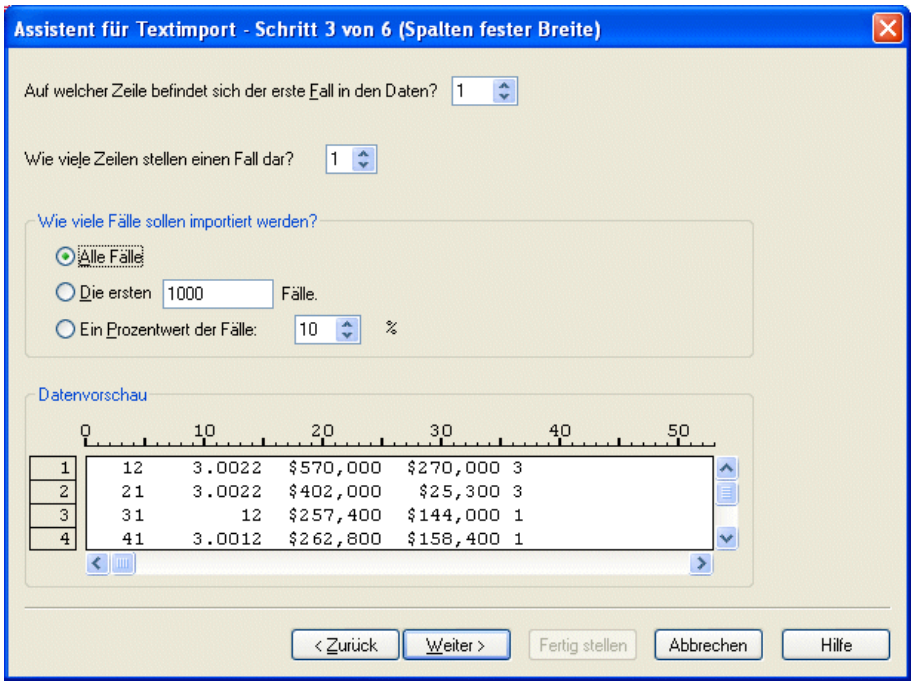
Wie sind die Fälle dargestellt? Gibt an, woran der Text-Assistent erkennt, wo jeder Fall endet und der nächste beginnt.

- **Jede Zeile stellt einen Fall dar.** Jede Zeile enthält nur einen Fall. Sehr häufig ist jeder Fall in einer einzigen Zeile enthalten. Hierbei kann es sich auch um eine sehr lange Zeile in einer Datendatei mit einer großen Anzahl von Variablen handeln. Wenn nicht alle Zeilen die gleiche Anzahl von Werten enthalten, wird die Anzahl der Variablen für jeden Fall durch die Zeile mit den meisten Datenwerten festgelegt. Fällen mit weniger Datenwerten werden fehlende Werte für die zusätzlichen Variablen zugewiesen.
- **Folgende Anzahl von Variablen stellt einen Fall dar.** Durch die festgelegte Anzahl von Variablen pro Fall wird dem Text-Assistenten mitgeteilt, an welcher Stelle eine Variable endet und die nächste beginnt. Eine Zeile kann mehrere Fälle enthalten, und Fälle können in der Mitte einer Zeile beginnen und in der nächsten Zeile fortgesetzt werden. Der Text-Assistent bestimmt das Ende jedes Falls unabhängig von der Anzahl der Zeilen anhand der Anzahl von eingelesenen Werten. Für jeden Fall müssen Datenwerte (oder durch Trennzeichen angezeigte fehlende Werte) für alle Variablen vorhanden sein. Ansonsten wird die Datendatei nicht ordnungsgemäß eingelesen.

Wie viele Fälle sollen importiert werden? Sie können alle Fälle der Datendatei, die ersten *n* Fälle (*n* ist eine von Ihnen festgelegte Zahl) oder eine Stichprobe mit einem bestimmten Prozentsatz der Fälle importieren. Da SPSS für jeden Fall eine unabhängige Pseudo-Zufallsentscheidung trifft, entspricht der Prozentsatz der tatsächlich ausgewählten Fälle nur ungefähr dem angegebenen Prozentwert. Je mehr Fälle sich in der Datendatei befinden, desto eher entspricht der Prozentsatz ausgewählter Fälle dem angegebenen Prozentsatz.

Text-Assistent: Schritt 3 (Dateien mit Spalten fester Breite)

Abbildung 3-16
Text-Assistent: Schritt 3 (für Dateien mit Spalten fester Breite)



In diesem Schritt erhalten Sie Informationen über die Fälle. Ein Fall ist vergleichbar mit einem Datensatz in einer Datenbank. So ist zum Beispiel jede Person, die einen Fragebogen ausfüllt, ein Fall.

Auf welcher Zeile befindet sich der erste Fall in den Daten? Gibt die erste Zeile der Datendatei an, die Datenwerte enthält. Wenn die erste(n) Zeile(n) der Datendatei aussagekräftige Beschreibungen oder anderen Text und somit keine Datenwerte enthalten, ist dies *nicht* Zeile 1.

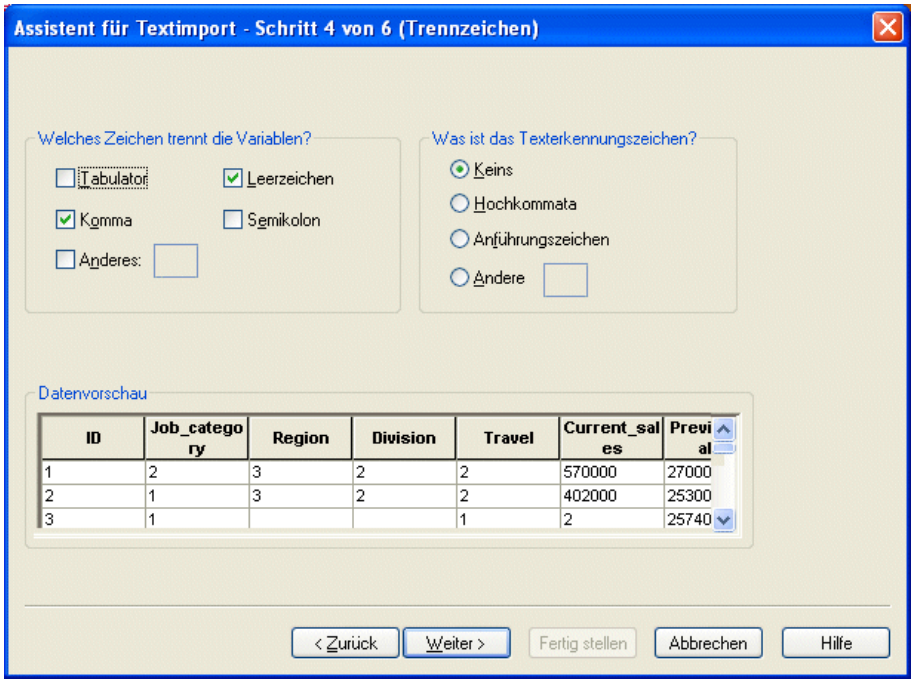
Wie viele Zeilen stellen einen Fall dar? Gibt an, woran der Text-Assistent erkennt, wo jeder Fall endet und der nächste beginnt. Jede Variable ist durch ihre Zeilennummer innerhalb des Falls und durch ihre Spaltenposition festgelegt. Sie müssen die Anzahl der Zeilen pro Fall angeben, damit die Daten richtig eingelesen werden können.

Wie viele Fälle sollen importiert werden? Sie können alle Fälle der Datendatei, die ersten *n* Fälle (*n* ist eine von Ihnen festgelegte Zahl) oder eine Stichprobe mit einem bestimmten Prozentsatz der Fälle importieren. Da SPSS für jeden Fall eine unabhängige Pseudo-Zufallsentscheidung trifft, entspricht der Prozentsatz der tatsächlich ausgewählten Fälle nur ungefähr dem angegebenen

Prozentwert. Je mehr Fälle sich in der Datendatei befinden, desto eher entspricht der Prozentsatz ausgewählter Fälle dem angegebenen Prozentsatz.

Text-Assistent: Schritt 4 (durch Trennzeichen getrennte Dateien)

Abbildung 3-17
Text-Assistent: Schritt 4 (für durch Trennzeichen getrennte Dateien)



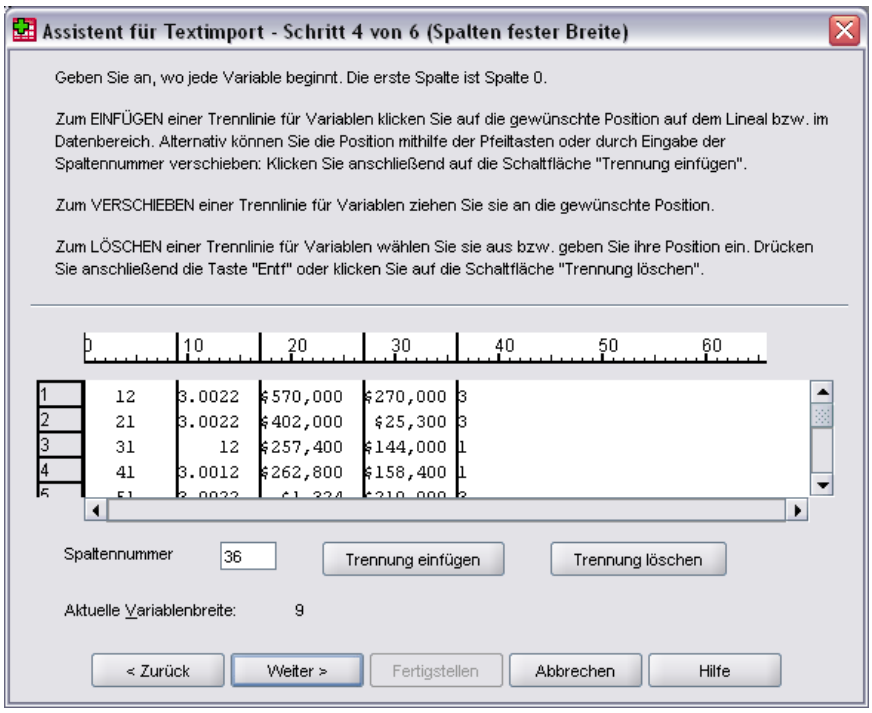
In diesem Schritt zeigt der Text-Assistent einen Vorschlag an, wie die Variablen aus der Datendatei eingelesen werden. Sie können diesen Vorschlag ändern.

Welches Zeichen trennt die Variablen? Geben Sie hier das Zeichen an, mit dem die Datenwerte getrennt werden. Sie können eine beliebige Kombination aus Leerzeichen, Kommata, Semikola, Tabulatoren und anderen Zeichen wählen. Mehrere aufeinander folgende Trennzeichen, die keine Datenwert einschließen, werden als fehlende Werte behandelt.

Was ist ein Texterkennungszeichen? Hierbei handelt es sich um Zeichen, mit denen Werte eingeschlossen werden, die Trennzeichen enthalten. Wenn beispielsweise ein Komma das Trennzeichen ist, werden Werte mit Kommata falsch eingelesen, sofern es keine Texterkennungszeichen gibt, die den Wert einschließen und damit verhindern, dass die Kommata in dem Wert als Trennzeichen zwischen Werten interpretiert werden. Bei Datendateien im CSV-Format, die aus Excel exportiert werden, werden Anführungszeichen (“) als Texterkennungszeichen verwendet. Die Texterkennungszeichen erscheinen am Anfang und am Ende des Werts, umschließen also den ganzen Wert.

Text-Assistent: Schritt 4 (Dateien mit Spalten fester Breite)

Abbildung 3-18
Text-Assistent: Schritt 4 (für Dateien mit Spalten fester Breite)



In diesem Schritt zeigt der Text-Assistent einen Vorschlag an, wie die Variablen aus der Datendatei eingelesen werden. Sie können diesen Vorschlag ändern. Vertikale Linien im Vorschaufenster kennzeichnen die Positionen, von denen der Text-Assistent annimmt, dass sie den Anfang der einzelnen Variablen in der Datei kennzeichnen.

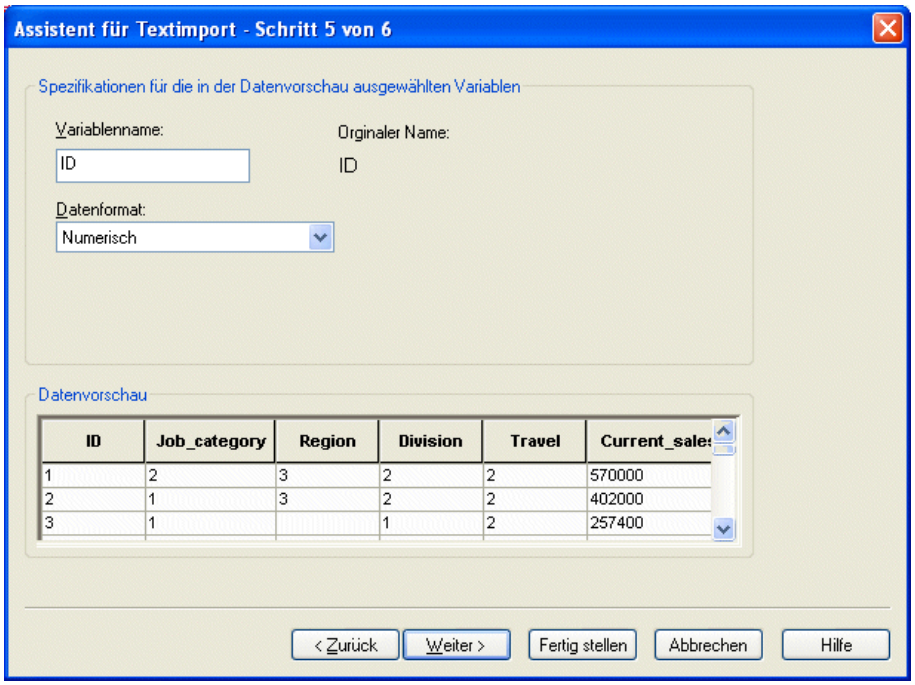
Fügen Sie erforderlichenfalls weitere Variablentrennlinien ein oder verschieben bzw. löschen Sie vorhandene Linien. Wenn für jeden Fall mehrere Zeilen verwendet werden, werden die Daten jeweils als eine einzige Zeile für jeden Fall angezeigt, wobei die nachfolgenden Zeilen am Ende der angezeigten Zeile angehängt werden.

Anmerkungen:

Bei computergenerierten Datendateien mit einem kontinuierlichen Fluß an Datenwerten ohne trennende Leerzeichen oder andere Zeichen kann es große Schwierigkeiten bereiten, zu bestimmen, wo die einzelnen Variablen beginnen. Bei dieser Art von Datendatei sind normalerweise Datendefinitionsdateien oder andere schriftliche Beschreibungen notwendig, in denen die Zeile und die Spalte für jede Variable aufgeführt sind.

Text-Assistent: Schritt 5

Abbildung 3-19
Text-Assistent: Schritt 5



In diesem Schritt legen Sie fest, welchen Variablennamen und welches Datenformat der Text-Assistent beim Einlesen der Variablen verwendet und welche Variablen in der endgültigen Datendatei enthalten sein werden.

Variablenname. Sie können die vom System vorgegebenen Variablennamen durch Ihre eigenen überschreiben. Beim Einlesen von Variablennamen aus der Datendatei ändert der Text-Assistent automatisch Variablennamen, die nicht den Regeln für Variablennamen entsprechen. Wählen Sie im Vorschaufenster eine Variable und geben Sie einen Variablennamen ein.

Datenformat. Wählen Sie im Vorschaufenster eine Variable und wählen Sie ein Format aus der Dropdown-Liste aus. Halten Sie zum Auswählen mehrerer aufeinander folgender Variablen beim Klicken die UMSCHALTTASTE gedrückt. Halten Sie beim Auswählen von mehreren nicht aufeinander folgenden Variablen beim Klicken die STRG-Taste gedrückt.

Formatoptionen im Text-Assistenten

Beim Einlesen von Variablen mit dem Text-Assistenten sind die folgenden Formatoptionen verfügbar:

Nicht importieren. Die gewählte(n) Variable(n) in der Datendatei wird/werden übersprungen.

Numerisch. Gültige Werte sind Ziffern, ein führendes Plus- oder Minuszeichen und ein Dezimaltrennzeichen.

String. Gültige Werte hierfür sind fast alle Zeichen, die auf der Tastatur eingegeben werden können, und eingebettete Leerzeichen. Bei Dateien mit Trennzeichen können Sie die Anzahl der Zeichen im Wert angeben, bis zu maximal 32,767. In der Standardeinstellung setzt der Text-Assistent die Zeichenanzahl auf den längsten String, der für die gewählte(n) Variable(n) gefunden werden kann. Bei Dateien mit Spalten fester Breite wird die Zeichenanzahl der Strings durch die Anordnung der Variablentrennlinien in Schritt 4 bestimmt.

Datum/Uhrzeit. Zu den gültigen Werten zählen Daten im allgemeinen Format *tt-mm-jjjj*, *mm/tt/jjjj*, *tt.mm.jjjj*, *jjjj/mm/tt*, *hh:mm:ss* und eine Vielzahl anderer Formate für das Datum und die Uhrzeit. Monate können durch arabische oder römische Ziffern und dreibuchstabige Abkürzungen dargestellt oder vollständig ausgeschrieben werden. Wählen Sie ein Datumsformat aus der Liste aus.

Dollar. Als zulässige Werte gelten Ziffern mit optionalem führenden Dollarzeichen und optionalen Kommata als Tausendertrennzeichen.

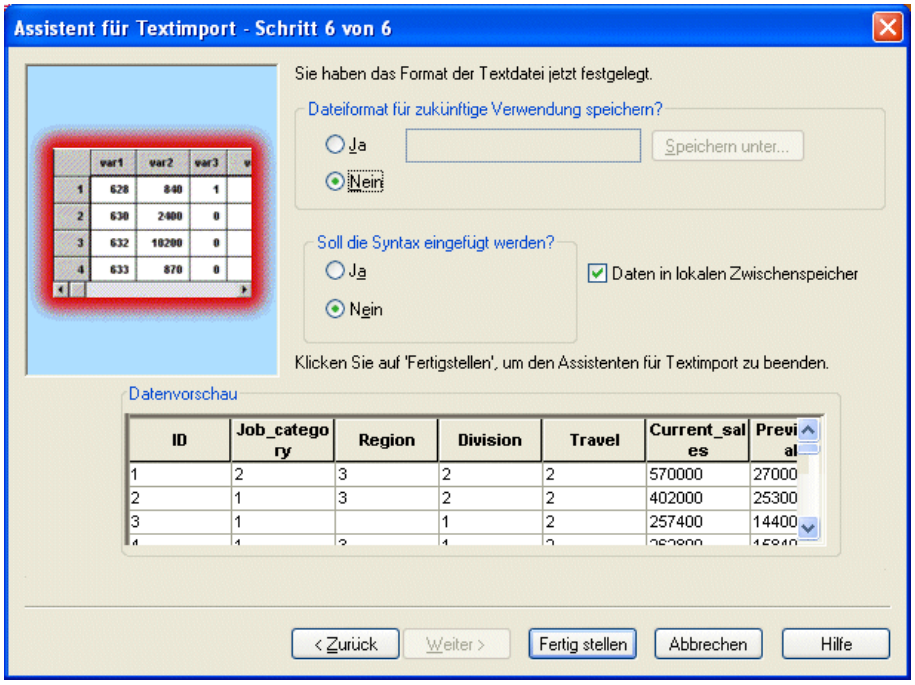
Komma. Als gültige Werte hierfür gelten Zahlen, in denen Dezimalstellen durch einen Punkt wiedergegeben und Kommata als Tausendertrennzeichen verwendet werden.

Punkt. Als gültige Werte hierfür gelten Zahlen, in denen Dezimalstellen durch ein Komma wiedergegeben und Punkte als Tausendertrennzeichen verwendet werden.

Anmerkung: Werte, die unzulässige Zeichen für das gewählte Format enthalten, werden als fehlende Werte behandelt. Werte, in denen eines der angegebenen Trennzeichen enthalten ist, werden als Mehrfachwerte behandelt.

Text-Assistent: Schritt 6

Abbildung 3-20
Text-Assistent: Schritt 6



Dies ist der letzte Schritt im Text-Assistenten. Sie können Ihre Einstellungen in einer Datei speichern, um sie beim Importieren ähnlicher Textdatendateien verwenden zu können. Sie können auch die vom Text-Assistenten erzeugte Syntax in ein Syntax-Fenster einfügen. Sie können die Syntax dann anpassen und/oder speichern, um sie bei anderen Sitzungen oder Produktionsjobs einsetzen zu können.

Daten in lokalen Zwischenspeicher. Ein Zwischenspeicher (Cache) für die Daten ist eine vollständige Kopie der Datendatei, die temporär auf der Festplatte gespeichert wird. Zwischenspeichern der Datendatei kann die Leistung verbessern.

Einlesen von Daten aus Dimensions

Unter Microsoft Windows-Betriebssytemen können Daten aus SPSS Dimensions-Produkten wie Quanvert, Quancept und mrInterview einlesen. (*Anmerkung*: Diese Funktion ist nur verfügbar, wenn SPSS unter Microsoft Windows-Betriebssytemen installiert ist.)

Um Dimensions-Datenquellen lesen zu können, müssen folgende Elemente installiert sein:

- .NET Framework
- Dimensions-Datenmodell und OLE DB Access

Die Versionen dieser Komponenten, die mit dieser Version kompatibel sind, können von der Installations-CD installiert werden und sind über das AutoPlay-Menü verfügbar. Dimensions-Datenquellen können nur im Modus für lokale Analysen eingelesen werden. Diese Funktion ist im Modus für verteilte Analyse mit dem SPSS-Server nicht verfügbar.

So lesen Sie Daten aus einer Dimensions-Datenquelle ein:

▶ Wählen Sie in einem beliebigen SPSS-Fenster die folgenden Befehle aus den Menüs aus:
Datei
Dimensionsdaten öffnen

▶ Geben Sie im Dialogfeld “Eigenschaften der Datenverknüpfung” auf der Registerkarte “Verbindung” die Metadatendatei, den Falldatentyp und die Falldatendatei an.

▶ Klicken Sie auf OK.

▶ Wählen Sie im Dialogfeld “Daten aus Dimensions importieren” die gewünschten Variablen und ggf. Fallauswahlkriterien aus.

▶ Klicken Sie auf OK, um die Daten einzulesen.

Registerkarte “Verbindung” im Dialogfeld “Eigenschaften der Datenverknüpfung”

Wenn Sie Daten aus einer Dimensions-Datenquelle einlesen möchten, müssen Sie Folgendes angeben:

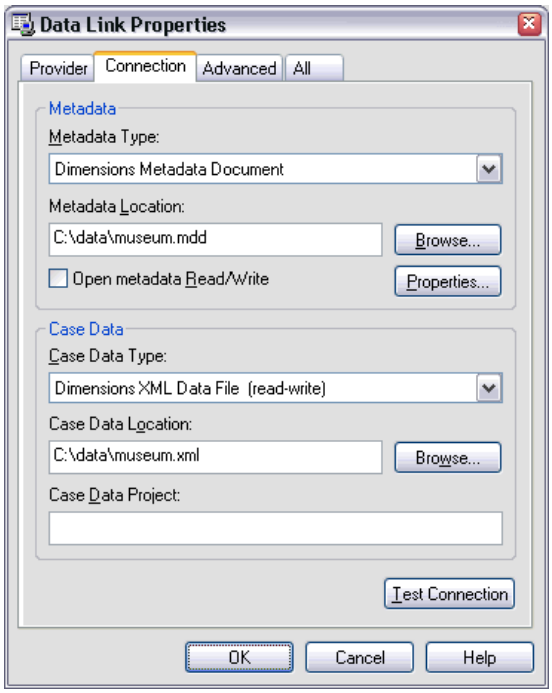
Speicherort der Metadaten. Dies ist die Metadaten-Dokumentdatei (*.mdd*), die Informationen zur Definition der Umfrage enthält.

Falldatentyp. Hierbei handelt es sich um das Format der Falldatendatei. Folgende Formate sind verfügbar:

- Quancept-Datendatei (DRS). Falldaten in einer Quancept-Datei im Format *.drs*, *.drz* oder *.dru*.
- Quanvert-Datenbank. Falldaten in einer Quanvert-Datenbank.
- Dimensions-Datenbank (MS SQL Server). Falldaten in einer relationalen SPSS MR-Datenbank auf einem SQL Server. Diese Option kann verwendet werden, um Daten einzulesen, die mit mrInterview erfasst wurden.
- Dimensions XML-Datendatei. Falldaten in einer XML-Datei.

Speicherort der Falldaten. Dies ist die Datei, die die Falldaten enthält. Das Format dieser Datei muss dem ausgewählten Falldatentyp entsprechen.

Abbildung 3-21
Eigenschaften der Datenverbindung: Registerkarte "Connection"



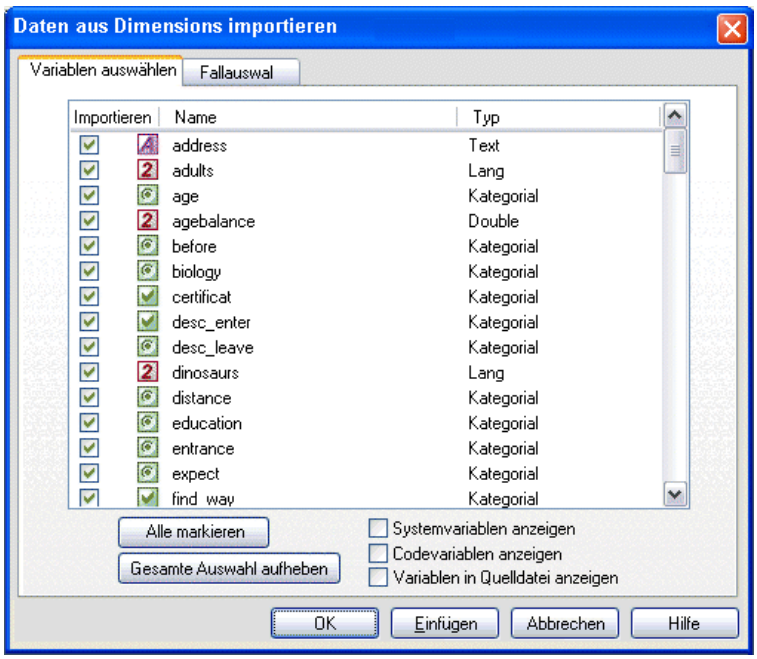
Anmerkung: Das Ausmaß, in welchem die weiteren Einstellungen auf der Registerkarte "Verbindung" sowie auf den weiteren Registerkarten im Dialogfeld "Eigenschaften der Datenverknüpfung" das Einlesen von Daten aus Dimensions in SPSS beeinträchtigen, ist nicht bekannt. Deshalb empfehlen wir, dass Sie keine dieser Einstellungen ändern.

Registerkarte "Variablen auswählen"

Sie können eine Untergruppe von einzulesenden Variablen auswählen. In der Standardeinstellung werden alle Standardvariablen in der Datenquelle angezeigt und ausgewählt.

- **Systemvariablen anzeigen.** Hiermit werden beliebige "Systemvariablen" angezeigt, einschließlich der Variablen, die den Interviewstatus kennzeichnen (*in progress*, *completed*. *finish date* usw.). Sie können dann die gewünschten Systemvariablen auswählen. In der Standardeinstellung werden alle Systemvariablen ausgeschlossen.
- **Codevariablen anzeigen.** Hiermit werden alle Variablen angezeigt, die die Codes darstellen, die für "weitere" Antworten mit offenen Ende für kategoriale Variablen verwendet werden. Sie können dann die gewünschten Codevariablen auswählen. In der Standardeinstellung werden alle Codevariablen ausgeschlossen.
- **Variablen in Quelldatei anzeigen.** Hiermit werden alle Variablen angezeigt, die die Dateinamen der Bilder von eingescannten Antworten enthalten. Sie können dann die gewünschten Variablen der Quelldatei auswählen. In der Standardeinstellung werden alle Variablen der Quelldatei ausgeschlossen.

Abbildung 3-22
Daten aus Dimensions importieren: Registerkarte "Variablen auswählen"



Registerkarte "Fallauswahl"

Bei Dimensions-Datenquellen, die Systemvariablen enthalten, können Sie Fälle auf der Basis einer Anzahl von Kriterien für Systemvariablen auswählen Sie müssen die entsprechenden Systemvariablen nicht in die Liste der einzulesenden Variablen aufnehmen. Die erforderlichen Systemvariablen müssen jedoch in den Quelldaten vorliegen, damit die Auswahlkriterien angewendet werden können. Wenn die erforderlichen Systemvariablen nicht in der Quelldatei vorhanden sind, werden die entsprechenden Auswahlkriterien ignoriert.

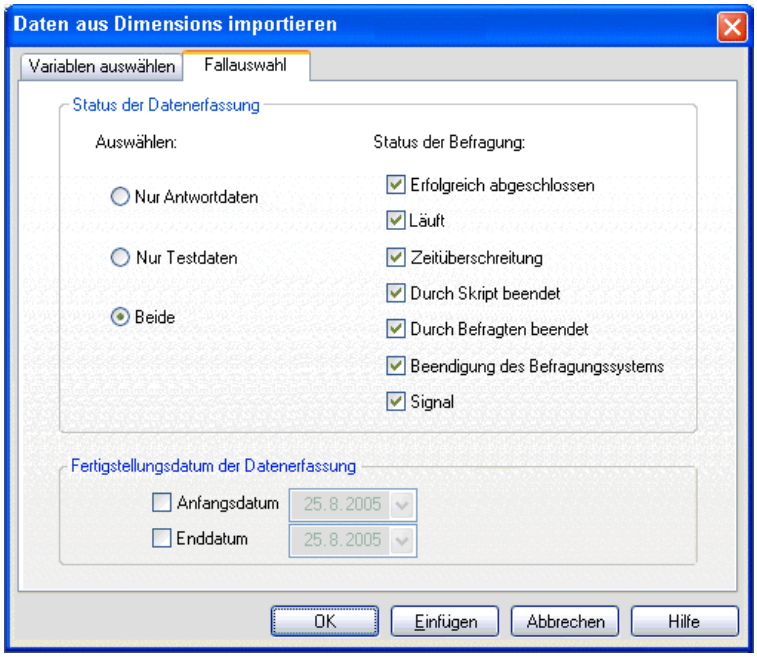
Status der Datenerfassung. Sie können Antwortdaten, Testdaten oder beide auswählen. Sie können Fälle auch auf der Basis einer beliebigen Kombination der folgenden Statusparameter für Interviews auswählen:

- Erfolgreich abgeschlossen
- Läuft
- Zeitüberschreitung
- Durch Skript beendet
- Durch Befragten beendet
- Beendigung des Befragungssystems
- Signal (durch eine Signalanweisung im Skript beendet)

Fertigstellungsdatum der Datenerfassung. Sie können Fälle auf der Basis des Fertigstellungsdatums der Datenerfassung auswählen.

- **Anfangsdatum.** Es werden Fälle berücksichtigt, bei denen die Datenerfassung am oder nach dem angegebenen Datum abgeschlossen wurde.
- **Enddatum.** Es werden Fälle berücksichtigt, bei denen die Datenerfassung vor dem angegebenen Datum abgeschlossen wurde. Hierbei werden *keine* Fälle berücksichtigt, für die die Datenerfassung am Enddatum abgeschlossen wurde.
- Wenn Sie das Anfangs- als auch das Enddatum angeben, ergibt dies einen Datumsbereich vom Anfangsdatum bis zum Tag vor dem Enddatum.

Abbildung 3-23
Daten aus Dimensions importieren: Registerkarte "Fallauswahl"



Informationen zur Datei

Eine SPSS-Datendatei enthält mehr als nur Rohdaten. Sie enthält außerdem Informationen zu Definitionen von Variablen. Dies umfaßt die folgenden Informationen:

- Variablennamen
- Variablenformate
- Beschreibende Variablen- und Wertelabels

Diese Informationen werden im Datenlexikon einer SPSS-Datendatei gespeichert. Mit dem Daten-Editor können die Informationen zu Definitionen der Variablen eingesehen werden. Es ist außerdem möglich, alle Informationen aus dem Datenlexikon der Arbeitsdatei oder einer beliebigen Datendatei anzuzeigen.

So zeigen Sie Datendatei-Informationen an:

▶ Wählen Sie die folgenden Befehle aus den Menüs im Fenster "Daten-Editor" aus:
Datei
Datendatei-Informationen anzeigen

▶ Wählen Sie für die derzeit geöffnete Datei die Option Arbeitsdatei.

▶ Wählen Sie für andere Datendateien die Option Externe Datei und wählen Sie dann die Datendatei aus.

Die Informationen zur Datendatei werden im Viewer angezeigt.

Speichern von Datendateien

Neben der Möglichkeit, Datendateien im SPSS-Format zu speichern, haben Sie auch die Möglichkeit, Daten in verschiedenen externen Formaten zu speichern, darunter:

- Excel- und andere Tabellenkalkulationsformate
- Tabulator- und kommagetrennte (CSV) Textdateien
- SAS
- Stata
- Datenbanktabellen

So speichern Sie geänderte Datendateien:

▶ Aktivieren Sie den Daten-Editor. Klicken Sie dazu auf eine beliebige Stelle des entsprechenden Fensters.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Datei
Speichern

Die geänderten Daten werden gespeichert. Dabei wird die vorherige Version der Datei überschrieben.

Anmerkung: Eine im Unicode-Modus gespeicherte Datendatei kann nicht von Versionen von SPSS vor 16.0 gelesen werden. Um eine Unicode-Datendatei in einem Format zu speichern, das von früheren Versionen gelesen werden kann, müssen Sie die Datei im Codepage-Modus öffnen und erneut speichern. Die Datei wird in der Kodierung gespeichert, die dem aktuellen Gebietsschema entspricht. Es kann Datenverlust auftreten, wenn die Datei Zeichen enthält, die vom aktuellen Gebietsschema nicht erkannt werden. Informationen zum Umschalten zwischen Unicode-Modus und Codeseiten-Modus finden Sie unter Optionen: Allgemein auf S. 491.

Speichern von Datendateien in externen Formaten

- ▶ Aktivieren Sie den Daten-Editor. Klicken Sie dazu auf eine beliebige Stelle des entsprechenden Fensters.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Speichern unter…
- ▶ Wählen Sie aus der Dropdown-Liste einen Dateityp aus.
- ▶ Geben Sie einen Namen für die neue Datendatei ein.

So schreiben Sie die Variablennamen in die erste Zeile eines Tabellenkalkulationsblatts oder die erste Zeile einer Tabulator-getrennten Textdatei:

- ▶ Klicken Sie im Dialogfeld “Daten speichern unter” auf Variablennamen im Arbeitsblatt speichern.

So speichern Sie Wertelabels anstelle von Datenwerten in Excel-Dateien:

- ▶ Klicken Sie im Dialogfeld “Daten speichern unter” auf Sofern definiert, Wertelabels statt Datenwerte speichern.

So speichern Sie Wertelabels in einer SAS-Syntaxdatei (nur nach Auswahl eines SAS-Dateityps aktiv):

- ▶ Klicken Sie im Dialogfeld “Daten speichern unter” auf Wertelabels in einer SAS-Datei speichern.

Informationen zum Exportieren von Daten in Datenbanktabellen finden Sie unter Export in eine Datenbank auf S. 54.

Informationen zum Exportieren von Daten zur Verwendung in Dimensions-Anwendungen finden Sie unter Export nach Dimensions auf S. 66.

Speichern von Daten: Datendateitypen

Sie können Ihre Daten in den folgenden Formaten speichern:

SPSS (*.sav). SPSS-Format.

- Programmversionen vor Version 7.5 können keine Datendateien lesen, die im SPSS-Format gespeichert werden. Im Unicode-Modus gespeicherte Datendateien können nicht von SPSS-Versionen vor SPSS 16.0 gelesen werden. Für weitere Informationen siehe Optionen: Allgemein in Kapitel 45 auf S. 491.

- Bei der Verwendung von Datendateien mit Variablennamen mit mehr als 8 Byte in den Versionen 10.x oder 11.x werden eindeutige, 8 Byte umfassende Versionen der Variablen verwendet. Die ursprünglichen Variablennamen bleiben jedoch für die Verwendung in Version 12.0 oder höher erhalten. Bei Versionen vor 10.0 gehen die ursprünglichen langen Variablennamen beim Speichern der Datendatei verloren.
- Wenn Sie Datendateien mit String-Variablen mit mehr als 255 Byte in Versionen vor Version 13.0 verwenden, werden diese String-Variablen in mehrere String-Variablen mit je 255 Byte aufgeteilt.

Version 7.0 (*.sav). Format der Version 7.0. Datendateien, die im Format der Version 7.0 gespeichert wurden, können von Version 7.0 und früheren Versionen eingelesen werden. Sie enthalten jedoch keine Definitionen von Mehrfachantworten-Sets oder Informationen aus Data Entry für Windows.

SPSS/PC+ (*.sys). Dateien im Format von SPSS/PC+. Wenn die Datendatei mehr als 500 Variablen enthält, werden nur die ersten 500 gespeichert. Bei Variablen mit mehr als einem benutzerdefinierten fehlenden Wert werden die zusätzlichen benutzerdefinierten fehlenden Werte in den ersten benutzerdefinierten fehlenden Wert umkodiert.

SPSS portable (*.por). Dateien im portablen Format, die von anderen Versionen von SPSS und Versionen unter anderen Betriebssystemen eingelesen werden können. Variablennamen sind auf 8 Byte begrenzt und werden gegebenenfalls automatisch in eindeutige 8 Byte umfassende Namen konvertiert. In den meisten Fällen ist es nicht mehr erforderlich, Daten im portablen Format zu speichern, da Datendateien im SPSS-Format von der Plattform bzw. vom Betriebssystem unabhängig sein sollten. Im Unicode-Modus können Datendateien nicht im portablen Dateiformat gespeichert werden. Für weitere Informationen siehe Optionen: Allgemein in Kapitel 45 auf S. 491.

Tabulator-getrennt (*.dat). Textdateien, bei denen die Werte durch Tabulatoren getrennt sind. (*Anmerkung*: In Stringwerte eingebettete Tabulatorzeichen bleiben in der tabulatorgetrennten Datei als Tabulatorzeichen erhalten. Es wird nicht zwischen in Werte eingebettete Tabulatorzeichen und Tabulatorzeichen, die zum Trennen von Werten dienen, unterschieden.)

Kommagetrennte (*.csv). Textdateien, bei denen die Werte durch Kommas oder Strichpunkte getrennt sind. Wenn aktuell als Dezimaltrennzeichen von SPSS ein Punkt verwendet wird, werden die Werte durch Kommas getrennt. Wenn aktuell als Dezimaltrennzeichen ein Komma verwendet wird, werden die Werte durch Strichpunkte getrennt.

Festes ASCII (*.dat). Textdateien im festen Format. Hierbei werden für alle Variablen die Standard-Schreibformate verwendet. Zwischen den Feldern der Variablen befinden sich weder Tabulator- noch Leerzeichen.

Excel 2007 (*.xlsx). Arbeitsmappe im XLSX-Format von Microsoft Excel 2007. Die maximale Anzahl an Variablen beträgt 16.000; alle zusätzlichen Variablen nach den ersten 16.000 werden verworfen. Wenn das Daten-Set mehr als eine Million Fälle enthält, werden in der Arbeitsmappe mehrere Arbeitsblätter erstellt.

Excel 97 bis 2003 (*.xls). Microsoft Excel 97-Arbeitsmappe. Die maximale Anzahl an Variablen beträgt 256; alle zusätzlichen Variablen nach den ersten 256 werden verworfen. Wenn das Daten-Set mehr als 65.356 Fälle enthält, werden in der Arbeitsmappe mehrere Arbeitsblätter erstellt.

Excel 2.1 (*.xls). Tabellenkalkulationsdateien im Format von Microsoft Excel 2,1. Die Dateien dürfen höchstens 256 Variablen und 16.384 Zeilen enthalten.

1-2-3 Version 3.0 (*.wk3). Tabellenkalkulationsdateien im Format von Lotus 1-2-3, Version 3.0. Es können höchstens 256 Variablen gespeichert werden.

1-2-3 Version 2.0 (*.wk1). Tabellenkalkulationsdateien im Format von Lotus 1-2-3, Version 2.0. Es können höchstens 256 Variablen gespeichert werden.

1-2-3 Version 1.0 (*.wks). Tabellenkalkulationsdateien im Format von Lotus 1-2-3 Version 1A. Es können höchstens 256 Variablen gespeichert werden.

SYLK (*.slk). Dateien im "Symbolic Link"-Format für Tabellenkalkulationsdateien von Microsoft Excel und Multiplan. Es können höchstens 256 Variablen gespeichert werden.

dBASE IV (*.dbf). dBASE IV-Format.

dBASE III (*.dbf). dBASE III-Format.

dBASE II (*.dbf). dBASE II-Format.

SAS v7+ für Windows, kurze Erweiterung (*.sd7). SAS Versionen 7–8 für Windows, kurzes Dateinamensformat.

SAS v7+ für Windows, lange Erweiterung (*.sas7bdat). SAS Versionen 7–8 für Windows, langes Dateinamensformat.

SAS v7+ für UNIX (*.ssd01). SAS v8 für UNIX.

SAS v6 für Windows (*.sd2). Dateien im Format SAS V6 für Windows/OS2.

SAS v6 für UNIX (*.ssd01). Dateien im Format SAS V6 für UNIX (Sun, HP, IBM).

SAS v6 für Alpha/OSF (*.ssd04). Dateiformat SAS V 6 für Alpha/OSF (DEC UNIX).

SAS Transport (*.xpt). SAS-Transportdatei.

Stata Version 8 Intercooled (*.dta).

Stata Version 8 SE (*.dta).

Stata Version 7 Intercooled (*.dta).

Stata Version 7 SE (*.dta).

Stata Version 6 (*.dta).

Stata Versionen 4–5 (*.dta).

Datei speichern: Optionen

Bei Tabellenkalkulationsdateien, tabulatorgetrennten und kommagetrennten Dateien können die Variablennamen in die erste Zeile der Datei geschrieben werden.

Speichern von Datendateien im Excel-Format

Daten können in drei verschiedenen Microsoft Excel-Dateiformaten gespeichert werden. Excel 2.1, Excel 97 und Excel 2007.

- Für Excel 2.1 und Excel 97 gilt eine Beschränkung auf 256 Spalten, daher werden nur die ersten 256 Variablen aufgenommen.
- Für Excel 2007 gilt eine Beschränkung auf 16.000 Spalten, daher werden nur die ersten 16.000 Variablen aufgenommen.
- Für Excel 2.1 gilt eine Beschränkung auf 16.384 Zeilen, daher werden nur die ersten 16.384 Fälle aufgenommen.
- Für Excel 97 und Excel 2007 gibt es außerdem Obergrenzen für die Anzahl der Zeilen pro Arbeitsblatt, Arbeitsmappen können jedoch aus mehreren Blättern bestehen und es werden mehrere Blätter erstellt, wenn die Obergrenze für ein einzelnes Blatt überschritten wird.

Variablentypen

In der folgenden Tabelle werden die Variablentypen der SPSS-Originaldaten und ihrer Entsprechungen in den exportierten Excel-Daten dargestellt.

SPSS Variablentyp	**Excel-Datenformat.**
Numerisch	0.00; #,##0.00; ...
Komma	0.00; #,##0.00; ...
Dollar	$#,##0_); ...
Datum	t-mmm-jjjj
Zeit	hh:mm:ss
String	Allgemein

Speichern von Datendateien im SAS-Format

Beim Speichern von Daten als SAS-Datei werden verschiedene Aspekte der Daten besonders behandelt. Dazu gehören:

- Bestimmte Zeichen, die für Variablennamen in SPSS zulässig sind, sind in SAS nicht gültig, beispielsweise @, # und $. Diese ungültigen Zeichen werden beim Exportieren der Daten durch einen Unterstrich ersetzt.
- SPSS-Variablennamen, die Mehrbyte-Zeichen enthalten (z. B. japanische oder chinesische Zeichen) werden in Variablenamen der allgemeinen Form *Vnnn* konvertiert, wobei *nnn* ein ganzzahliger Wert ist.
- SPSS-Variablenlabels mit mehr als 40 Zeichen werden beim Exportieren in eine SAS v6-Datei abgeschnitten.
- Sofern SPSS-Variablenlabels vorhanden sind, werden sie den entsprechenden SAS-Variablenlabels zugeordnet. Wenn die SPSS-Daten keine Variablenlabels enthalten, wird dem SAS-Variablenlabel der Variablenname zugeordnet.
- Während in SPSS zahlreiche systemdefinierte fehlende Werte zulässig sind, kann es in SAS nur einen einzigen systemdefiniert fehlenden Wert geben. Daher werden alle systemdefiniert fehlenden Werte in SPSS nur einem systemdefinierten fehlenden Wert in der SAS-Datei zugeordnet.

Speichern von Wertelabels

Sie haben die Möglichkeit, die der Datendatei zugeordneten Werte und Wertelabels in einer SAS-Syntaxdatei zu speichern. Wenn beispielsweise die Wertelabels für die Datendatei *cars.sav* exportiert werden, enthält die erzeugte Syntaxdatei folgende Zeilen:

```
libname library '\spss\' ;

proc format library = library ;
   value ORIGIN /* Herstellungsland */
      1 = 'Amerika'
      2 = 'Europa'
      3 = 'Japan' ;
   value CYLINDER /* Anzahl der Zylinder */
      3 = '3 Zylinder'
      4 = '4 Zylinder'
      5 = '5 Zylinder'
      6 = '6 Zylinder'
      8 = '8 Zylinder'
   value FILTER__ /* zylinder = 1 | zylinder = 2 (FILTER) */
      0 = 'Nicht ausgewählt'
      1 = 'Ausgewählt' ;

proc datasets library = library ;
modify cars;
   format    ORIGIN ORIGIN.;
   format  CYLINDER CYLINDER.;
   format  FILTER__ FILTER__.;
quit;
```

Diese Funktion wird für SAS-Transportdateien nicht unterstützt.

Variablentypen

In der folgenden Tabelle werden die Variablentypen der SPSS-Originaldaten und ihrer Entsprechungen in den exportierten SAS-Daten dargestellt.

SPSS Variablentyp	SAS-Variablentyp	SAS-Datenformat
Numerisch	Numerisch	12
Komma	Numerisch	12
Punkt	Numerisch	12
Wissenschaftliche Notation	Numerisch	12

SPSS Variablentyp	SAS-Variablentyp	SAS-Datenformat
Datum	Numerisch	(Datum), z. B. MMDDYY10, ...
Datum (Uhrzeit)	Numerisch	Time18
Dollar	Numerisch	12
Spezielle Währung	Numerisch	12
String	Zeichen	$8

Speichern von Datendateien im Stata-Format

- Die Daten können im Format der Stata Versionen 5–8 geschrieben werden, sowohl im Intercooled- als auch im SE-Format (nur Versionen 7 und 8).
- Die Datendateien, die im Format von Stata 5 gespeichert wurden, können von Stata 4 eingelesen werden.
- Die ersten 80 Byte von Variablenlabels werden als Stata-Variablenlabels gespeichert.
- Bei numerischen Variablen werden die ersten 80 Byte der Variablenlabels als Stata-Wertelabels gespeichert. Bei String-Variablen werden die Wertelabels verworfen.
- Beim Format der Versionen 7 und 8 werden die ersten 32 Byte der Variablennamen unter Berücksichtigung der Groß- und Kleinschreibung als Stata-Variablennamen gespeichert. Beim Format von früheren Versionen werden die ersten 8 Byte der Variablennamen als Stata-Variablennamen gespeichert. Alle Zeichen außer Buchstaben, Ziffern und Unterstrichen werden in Unterstriche umgewandelt.
- SPSS-Variablennamen, die Mehrbyte-Zeichen enthalten (z. B. japanische oder chinesische Zeichen) werden in Variablenamen der allgemeinen Form *Vnnn* konvertiert, wobei *nnn* ein ganzzahliger Wert ist.
- Beim Format der Versionen 5–6 und den Intercooled-Versionen 7–8 werden die ersten 80 Byte der String-Werte gespeichert. Beim Format von Stata SE 7–8 werden die ersten 244 Byte der String-Werte gespeichert.
- Beim Format der Versionen 5–6 und den Intercooled-Versionen 7–8 werden nur die ersten 2.047 Variablen gespeichert. Beim Format von Stata SE 7–8 werden nur die ersten 32.767 Variablen gespeichert.

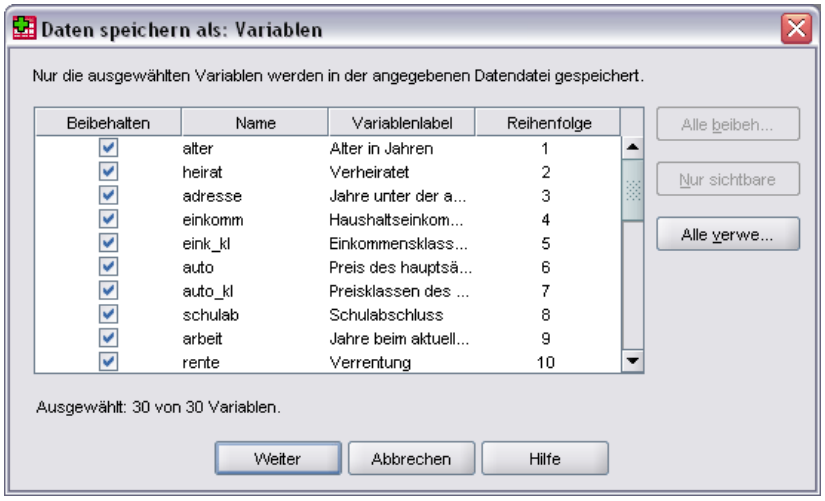
SPSS Variablentyp	Stata-Variablentyp	Stata-Datenformat
Numerisch	Numerisch	g
Komma	Numerisch	g
Punkt	Numerisch	g
Wissenschaftliche Notation	Numerisch	g
Datum*, Datum/Zeit	Numerisch	D_m_Y
Zeit, DTime	Numerisch	g (Anzahl der Sekunden)
Wkday	Numerisch	g (1–7)
Monat	Numerisch	g (1–12)
Dollar	Numerisch	g

SPSS Variablentyp	Stata-Variablentyp	Stata-Datenformat
Spezielle Währung	Numerisch	g
String	String	s

*Date, Adate, Edate, SDate, Jdate, Qyr, Moyr, Wkyr

Speichern von Untergruppen von Variablen

Abbildung 3-24
Dialogfeld "Daten speichern als: Variablen"



Im Dialogfeld "Daten speichern als: Variablen" können Sie die Variablen auswählen, die in der neuen Datendatei gespeichert werden sollen. In der Standardeinstellung werden alle Variablen gespeichert. Heben Sie die Auswahl für die Variablen auf, die nicht gespeichert werden sollen, oder klicken Sie auf Alle verwerfen und wählen Sie dann die zu speichernden Variablen aus.

Nur sichtbare. Wählt nur Variablen in derzeit in Verwendung befindlichen Variablen-Sets aus. Für weitere Informationen siehe Variablen-Sets verwenden in Kapitel 44 auf S. 487.

So speichern Sie eine Untergruppe von Variablen:

- Aktivieren Sie den Daten-Editor. Klicken Sie dazu auf eine beliebige Stelle des entsprechenden Fensters.
- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Datei
  Speichern unter…
- Klicken Sie auf Variablen.
- Wählen Sie die Variablen aus, die Sie speichern möchten.

Export in eine Datenbank

Mit dem Assistenten für den Datenbank-Export haben Sie folgende Möglichkeiten:

- Ersetzen der Werte in bestehenden Datenbankfeldern (Spalten) oder Hinzufügen neuer Felder zu einer Tabelle.
- Anhängen neuer Datensätze (Zeilen) zu einer Datenbanktabelle.
- Vollständiger Austausch einer Datenbanktabelle oder Erstellen einer neuen Tabelle.

So exportieren Sie Daten in eine Datenbank:

▶ Wählen Sie aus den Menüs in dem Fenster des Daten-Editors für das Daten-Set, das die zu exportierenden Daten enthält, folgende Optionen aus:

Datei
  In Datenbank exportieren

▶ Wählen Sie die Datenbankquelle aus.

▶ Befolgen Sie die Anweisungen des Exportassistenten, um die Daten zu exportieren.

Erstellen von Datenbankfeldern aus SPSS-Variablen

Beim Erstellen neuer Felder (Hinzufügen von Feldern zu einer bestehenden Datenbanktabelle, Erstellen einer neuen Tabelle, Ersetzen einer Tabelle) können Sie Feldnamen, Datentyp und Breite (sofern anwendbar) angeben.

Feldname. Die Standardfeldnamen stimmen mit den SPSS-Variablennamen überein. Sie können die Feldnamen auf jeden Namen ändern, der im betreffenden Datenbankformat zulässig ist. So sind bei vielen Datenbanken bestimmte Zeichen in Feldnamen zulässig, die in Variablennamen nicht erlaubt sind, beispielsweise Leerzeichen. Daher kann ein Variablenname wie *Anruf_ausstehend* in den Feldnamen *Anruf ausstehend* geändert werden.

Typ. Der Exportassistent nimmt erste Datentypzuweisungen auf der Grundlage der standardmäßigen ODBC-Datentypen oder der Datentypen vor, die im ausgewählten Datenbankformat, das dem definierten SPSS-Datenformat am nächsten kommt, zulässig sind. Die Datenbanken können jedoch Typunterscheidungen vornehmen, für die es in SPSS keine direkte Entsprechung gibt und umgekehrt. So werden die meisten numerischen Werte in SPSS als Gleitkommawerte mit doppelter Genauigkeit gespeichert, wohingegen numerische Datentypen in Datenbanken Gleitkommazahlen (Float (Double)), Ganzzahlen, reelle Zahlen usw. sein können. Außerdem gibt es in vielen Datenbanken keine Entsprechungen zu SPSS-Zeitformaten. Sie können den Datentyp in jeden Datentyp ändern, der in der Dropdown-Liste zur Verfügung steht.

Im Allgemeinen sollte der Grunddatentyp (String oder numerisch) der Variablen mit dem Grunddatentyp des Datenbankfelds übereinstimmen. Wenn die Datentypen nicht übereinstimmen und dieses Problem nicht von der Datenbank behoben werden kann, führt dies zu einem Fehler und es werden keine Daten in die Datenbank exportiert. Wenn Sie beispielsweise eine Stringvariable in ein Datenbankfeld mit einem numerischen Datentyp exportieren, wird ein Fehler ausgegeben, wenn irgendwelche Werte der Stringvariablen nichtnumerische Zeichen enthalten.

Breite. Sie können die definierte Breite für die Felder vom Typ “String” (char, varchar) ändern. Die Breite numerischer Felder richtet sich nach dem Datentyp.

Standardmäßig sind die SPSS-Variablenformate anhand des folgenden allgemeinen Schemas bestimmten Feldtypen in der Datenbank zugeordnet. Die tatsächlichen Datenbankfeldtypen können je nach Datenbank variieren.

SPSS Variablenformat	Feldtyp der Datenbank
Numerisch	Float oder Double
Komma	Float oder Double
Punkt	Float oder Double
Wissenschaftliche Notation	Float oder Double
Datum	Date oder Datetime oder Timestamp
Datetime	Datetime oder Timestamp
Zeit, DTime	Float oder Double (Anzahl der Sekunden)
Wkday	Integer (1–7)
Monat	Integer (1–12)
Dollar	Float oder Double
Spezielle Währung	Float oder Double
String	Char oder Varchar

Benutzerdefiniert fehlende Werte

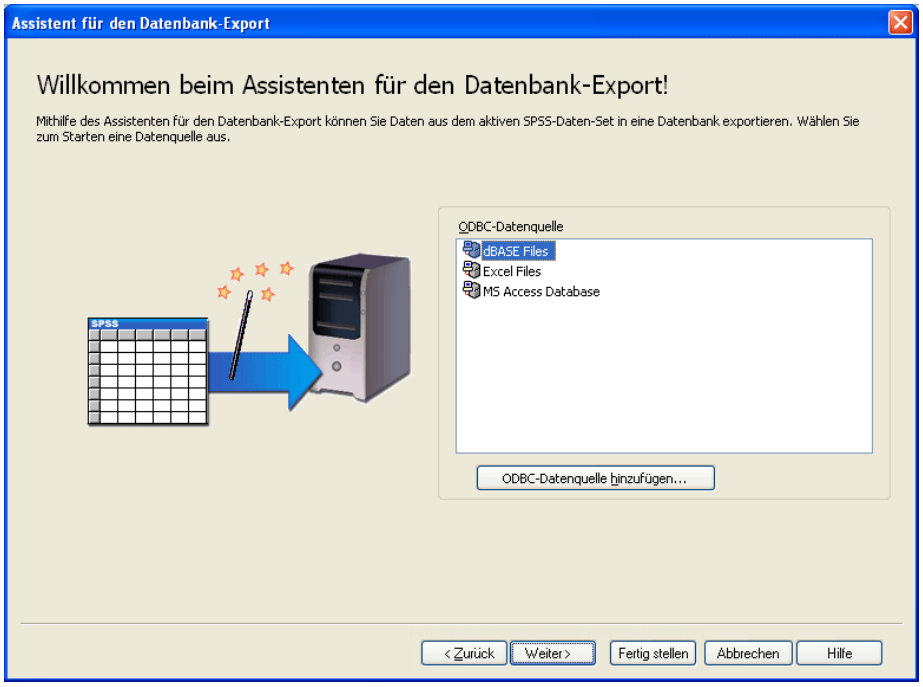
Es gibt zwei Optionen für den Umgang mit benutzerdefiniert fehlenden Werten beim Export von Daten aus Variablen in Datenbankfelder:

- **Als gültige Werte exportieren.** Benutzerdefiniert fehlende Werte werden als reguläre, gültige, nichtfehlende Werte behandelt.
- **Benutzerdefiniert fehlende numerische Werte als Nullen exportieren und benutzerdefiniert fehlende String-Werte als Leerzeichen exportieren.** Numerische benutzerdefiniert fehlende Werte werden wie systemdefiniert fehlende Werte behandelt. Benutzerdefiniert fehlende Stringwerte werden in Leerzeichen konvertiert (Strings können nicht systemdefiniert fehlend sein).

Auswählen einer Datenquelle

Im ersten Fenster des Assistenten für den Datenbank-Export wählen Sie die Datenquelle aus, in die die Daten exportiert werden sollen.

Abbildung 3-25
Assistent für den Datenbank-Export, Auswahl einer Datenquelle



Sie können Daten in jede Datenbankquelle exportieren, für die Sie über den entsprechenden ODBC-Treiber verfügen. (*Anmerkung*: Das Exportieren von Daten in OLE DB-Datenquellen wird nicht unterstützt.)

Wenn Sie noch keine ODBC-Datenquelle konfiguriert haben oder eine neue Datenquelle hinzufügen möchten, klicken Sie auf ODBC-Datenquelle hinzufügen.

- Bei Linux-Betriebssystemen ist diese Schaltfläche nicht verfügbar. ODBC-Datenquellen werden in *odbc.ini* angegeben und für die *ODBCINI*-Umgebungsvariablen muss der Speicherort der betreffenden Datei festgelegt sein. Weitere Informationen finden Sie in der Hilfe zu Ihren Datenbanktreibern.
- Im Modus für verteilte Analysen (verfügbar mit SPSS Server) steht diese Schaltfläche nicht zur Verfügung. Wenn Sie Datenquellen im Modus für verteilte Analysen hinzufügen möchten, wenden Sie sich an Ihren Systemadministrator.

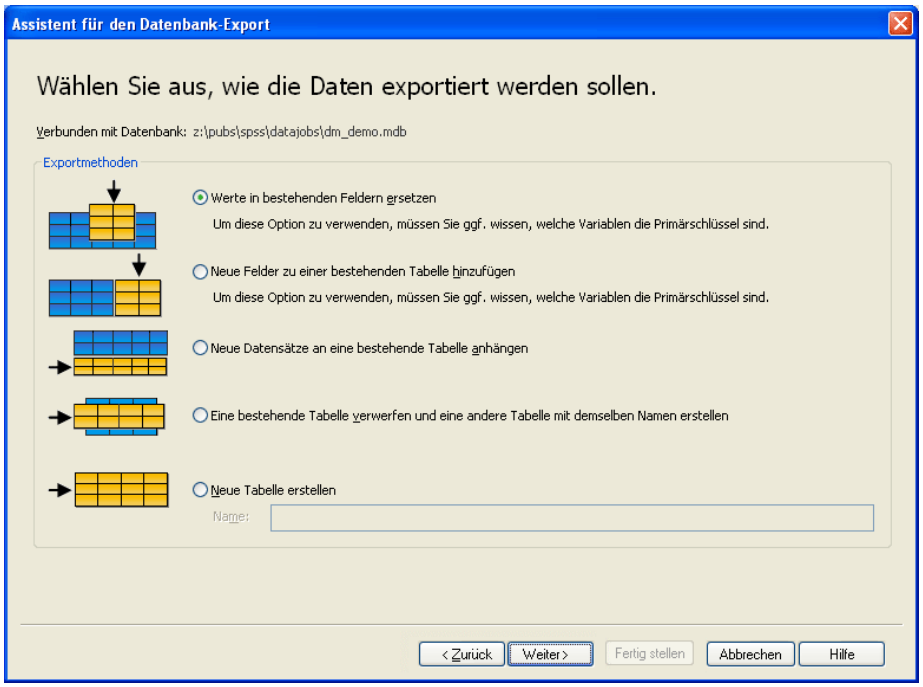
Eine ODBC-Datenquelle besteht aus zwei wichtigen Informationen: dem Treiber, der zum Zugreifen auf die Daten verwendet wird, und dem Speicherort der Datenbank, auf die Sie zugreifen möchten. Wenn Sie Datenquellen definieren möchten, muss der entsprechende Treiber installiert sein. Für den Modus für lokale Analysen können Sie Treiber für eine Vielzahl von Datenbankformaten von der SPSS-Installations-CD-ROM installieren.

Bei einigen Datenquellen benötigen Sie einen Anmeldenamen und ein Passwort, um mit dem nächsten Schritt fortzufahren.

Auswahl der Exportweise für die Daten

Nach der Auswahl der Datenquelle geben Sie an, wie die Daten exportiert werden sollen.

Abbildung 3-26
Assistent für den Datenbank-Export, Auswahl der Exportweise



Für den Export von Daten in eine Datenbank stehen folgende Optionen zur Verfügung:

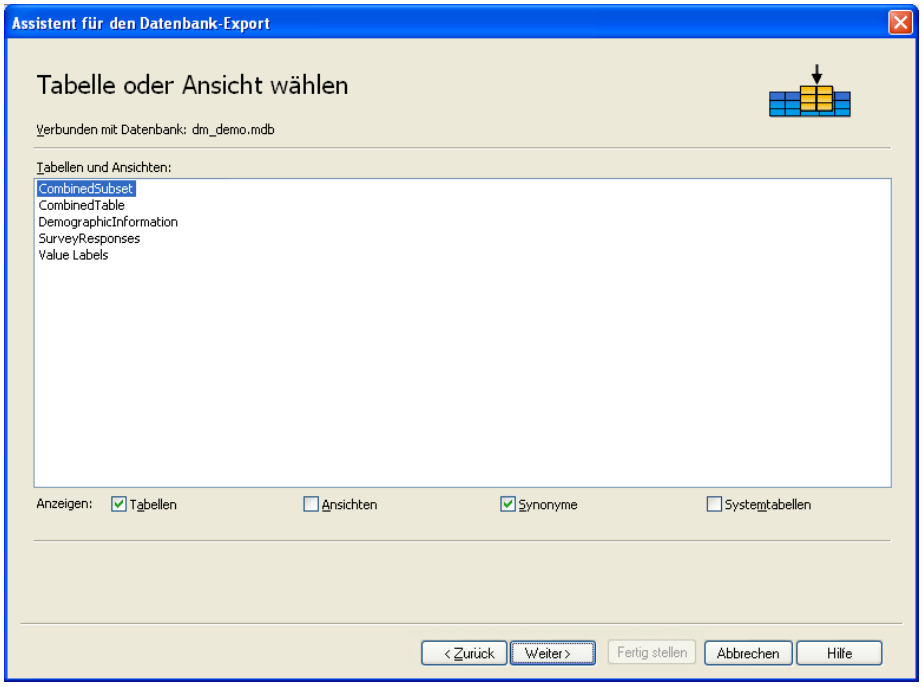
- **Werte in bestehenden Feldern ersetzen.** Ersetzt die Werte der ausgewählten Felder in einer bestehenden Tabelle mit Werten aus den ausgewählten Variablen im aktiven Daten-Set. Für weitere Informationen siehe Ersetzen von Werten in bestehenden Feldern auf S. 61.
- **Neue Felder zu einer bestehenden Tabelle hinzufügen.** Erstellt neue Felder in einer bestehenden Tabelle, die die Werte der ausgewählten Variablen im aktiven Daten-Set enthalten. Für weitere Informationen siehe Hinzufügen neuer Felder auf S. 62. Diese Option ist nicht für Excel-Dateien verfügbar.
- **Neue Datensätze an eine bestehende Tabelle anhängen.** Fügt neue Datensätze (Zeilen) zu einer bestehenden Datei hinzu, die die Werte aus Fällen in der Arbeitsdatei enthält. Für weitere Informationen siehe Anhängen neuer Datensätze (Fälle) auf S. 63.
- **Eine bestehende Tabelle verwerfen und eine andere Tabelle mit demselben Namen erstellen.** Löscht die angegebene Tabelle und erstellt eine neue Tabelle mit demselben Namen, die ausgewählte Variablen aus der Arbeitsdatei enthält. Alle Informationen aus der ursprünglichen Tabelle, einschließlich der Definitionen von Feldeigenschaften (z. B. Primärschlüssel, Datentypen), gehen verloren. Für weitere Informationen siehe Erstellen einer neuen Tabelle bzw. Ersetzen einer Tabelle auf S. 64.
- **Neue Tabelle erstellen.** Erstellt eine neue Tabelle in der Datenbank, die Daten aus ausgewählten Variablen aus der Arbeitsdatei enthält. Bei dem Namen kann es sich um jeden beliebigen Wert handeln, der gemäß der Datenquelle als Tabellenname zulässig ist. Es darf nicht der Name

einer bestehenden Tabelle oder Ansicht in der Datenbank verwendet werden. Für weitere Informationen siehe Erstellen einer neuen Tabelle bzw. Ersetzen einer Tabelle auf S. 64.

Auswählen einer Tabelle

Wenn Sie eine Tabelle in der Datenbank bearbeiten oder ersetzen möchten, müssen Sie diese zunächst auswählen. In diesem Fenster des Assistenten für den Datenbank-Export wird eine Liste der Tabellen und Ansichten in der ausgewählten Datenbank angezeigt.

Abbildung 3-27
Assistent für den Datenbank-Export, Auswahl einer Tabelle bzw. Ansicht



Standardmäßig werden in dieser Liste nur Standarddatenbanktabellen angezeigt. Sie können festlegen, welche Art von Elementen in der Liste angezeigt werden sollen:

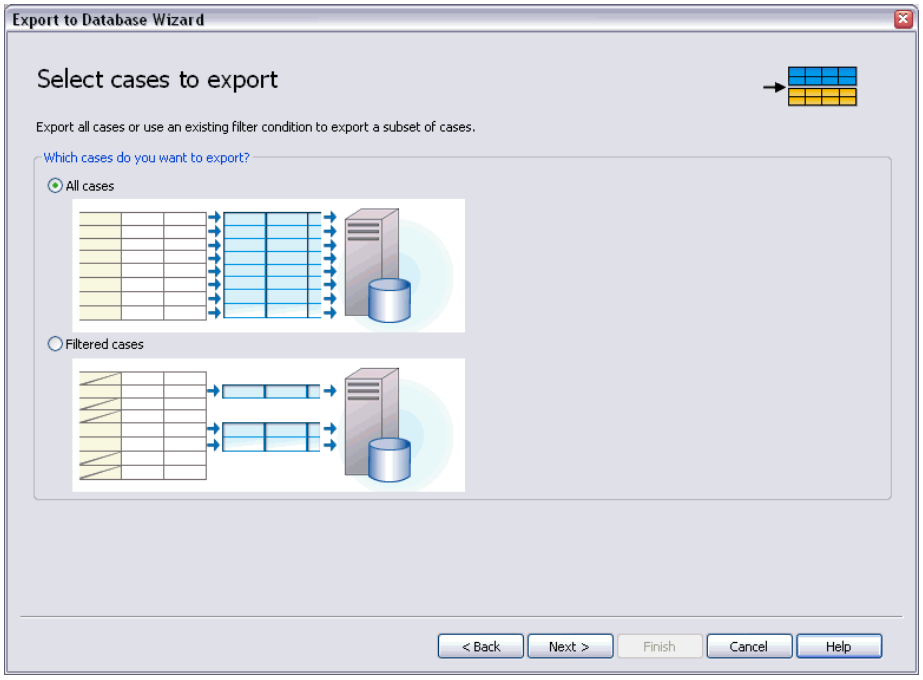
- **Tabellen.** Standarddatenbanktabellen.
- **Ansichten.** Ansichten sind virtuelle oder dynamische “Tabellen”, die durch Abfragen definiert werden. Dazu können Verknüpfungen mehrerer Tabellen und/oder Felder gehören, die aus Berechnungen auf der Grundlage der Werte in anderen Feldern abgeleitet wurden. Sie können Datensätze anhängen oder Werte bestehender Felder in Ansichten ersetzen, je nach Strukturierung der Ansicht können jedoch Beschränkungen hinsichtlich der bearbeitbaren Felder in Kraft sein. Es können beispielsweise keine abgeleiteten Felder bearbeitet, keine Felder zu einer Ansicht hinzugefügt und keine Ansichten ersetzt werden.

- **Synonyme.** Ein Synonym ist ein Alias für eine Tabelle oder eine Ansicht und wird normalerweise in einer Abfrage definiert.
- **Systemtabellen.** Systemtabellen definieren Datenbankeigenschaften. In einigen Fällen können Standarddatenbanktabellen als Systemtabellen klassifiziert sein und nur bei Auswahl dieser Option angezeigt werden. Der Zugriff auf eigentliche Systemtabellen ist häufig auf Datenbankadministratoren beschränkt.

Auswahl der zu exportierenden Fälle

Als Fallauswahl im Assistenten für den Datenbank-Export sind nur entweder alle Fälle oder die mithilfe einer zuvor definierten Filterbedingung ausgewählten Fälle möglich. Wenn keine Fallfilterung in Kraft ist, wird dieses Fenster nicht angezeigt und alle Fälle im aktiven Daten-Set werden exportiert.

Abbildung 3-28
Assistent für den Datenbank-Export, Auswahl der zu exportierenden Fälle



Weitere Informationen zur Definition einer Filterbedingung für die Fallauswahl finden Sie unter Fälle auswählen auf S. 200.

Abgleich zwischen Fällen und Datensätzen

Beim Hinzufügen von Feldern (Spalten) zu einer bestehenden Tabelle bzw. beim Ersetzen der Werte von bestehenden Feldern müssen Sie sicherstellen, dass jeder Fall (Zeile) im aktiven Daten-Set korrekt mit dem zugehörigen Datensatz in der Datenbank abgeglichen ist.

- In der Datenbank wird das Feld bzw. das Set von Feldern, das die einzelnen Datensätze eindeutig identifiziert, häufig als **Primärschlüssel** bezeichnet.

- Sie müssen angeben, welche Variable(n) dem/den Primärschlüsselfeld(ern) oder anderen Feldern entsprechen, die die einzelnen Datensätze eindeutig identifizieren.
- Bei den Feldern muss es sich nicht um den Primärschlüssel in der Datenbank handeln, doch der Feldwert bzw. die Kombination der Feldwerte muss für jeden Fall eindeutig sein.

So gleichen Sie die Variablen mit Feldern in der Datenbank ab, die die einzelnen Datensätze eindeutig identifizieren:

▶ Ziehen Sie die Variable(n) auf die entsprechenden Datenbankfelder und legen Sie sie dort ab.

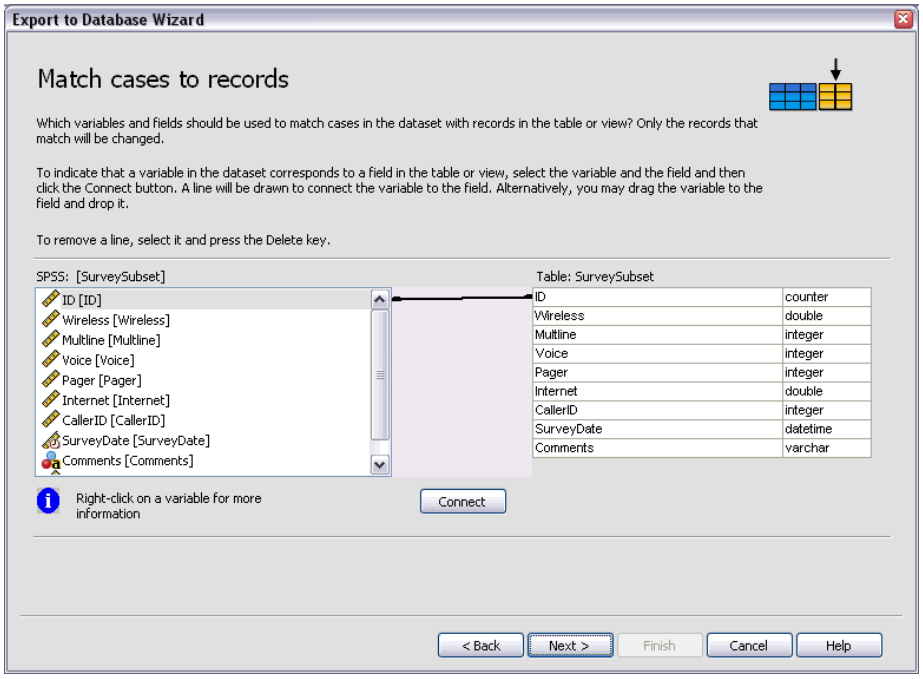
oder

▶ Wählen Sie eine Variable aus der Liste der Variablen, wählen Sie das entsprechende Feld in der Datenbanktabelle aus und klicken Sie auf Verbinden.

So löschen Sie eine Verbindungslinie:

▶ Wählen Sie die Verbindungslinie aus und drücken Sie die Entf-Taste.

Abbildung 3-29
Assistent für den Datenbank-Export, Abgleich zwischen Fällen und Datensätzen



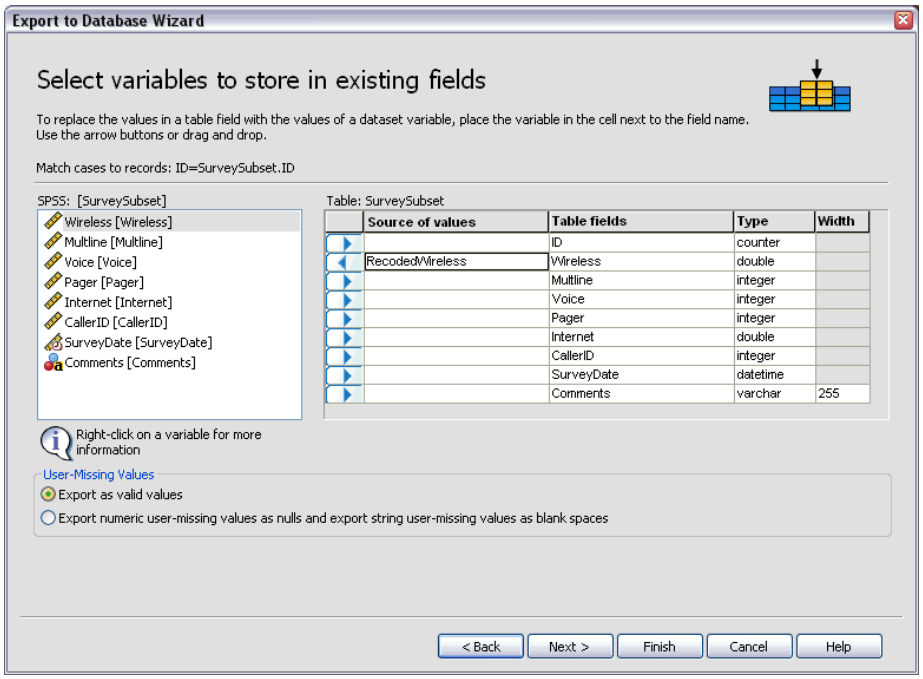
Anmerkung: Die SPSS-Variablennamen und die Feldnamen in der Datenbank sind nicht unbedingt identisch (da Datenbankfeldnamen Zeichen enthalten können, die in SPSS-Variablennamen nicht zulässig sind), doch wenn die Arbeitsdatei aus der von Ihnen bearbeiteten Datenbanktabelle erstellt wurde, sind die Variablennamen bzw. die Variablenlabels normalerweise den Feldnamen in der Datenbank zumindest ähnlich.

Ersetzen von Werten in bestehenden Feldern

So können Sie Werte von bestehenden Feldern in einer Datenbank ersetzen:

- Wählen Sie im Fenster Wählen Sie aus, wie die Daten exportiert werden sollen des Assistenten für den Datenbank-Export die Option Werte in bestehenden Feldern ersetzen aus.
- Wählen Sie im Feld Tabelle oder Ansicht wählen die Datenbanktabelle aus.
- Gleichen Sie im Feld Fälle mit Datensätzen abgleichen die Variablen, die die einzelnen Fälle eindeutig identifizieren, mit den entsprechenden Datenbankfeldnamen ab.
- Ziehen Sie für jedes Feld, für das Sie Werte ersetzen möchten, die Variable, die die neuen Werte enthält, in die Spalte Quelle der Werte neben dem entsprechenden Datenbankfeldnamen.

Abbildung 3-30
Assistent für den Datenbank-Export, Ersetzen von Werten bestehender Felder



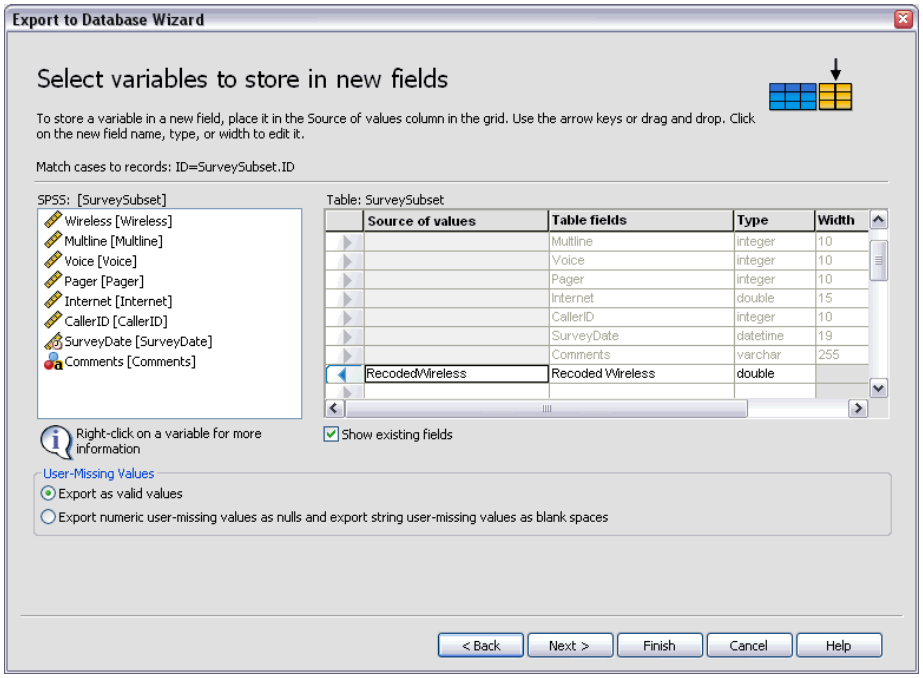
- Im Allgemeinen sollte der Grunddatentyp (String oder numerisch) der Variablen mit dem Grunddatentyp des Datenbankfelds übereinstimmen. Wenn die Datentypen nicht übereinstimmen und dieses Problem nicht von der Datenbank behoben werden kann, führt dies zu einem Fehler und es werden keine Daten in die Datenbank exportiert. Wenn Sie beispielsweise eine Stringvariable in ein Datenbankfeld mit einem numerischen Datentyp exportieren (z. B. double, reell, ganze Zahl), wird ein Fehler ausgegeben, wenn irgendwelche Werte der Stringvariablen nichtnumerische Zeichen enthalten. Der Buchstabe *a* in dem Symbol neben einer Variablen kennzeichnet eine String-Variable.
- Feldnamen, Typ und Breite können nicht bearbeitet werden. Die ursprünglichen Attribute des Datenbankfelds werden beibehalten; es werden lediglich die Werte ersetzt.

Hinzufügen neuer Felder

So fügen Sie neue Felder zu einer bestehenden Datenbanktabelle hinzu:

- Wählen Sie im Fenster Wählen Sie aus, wie die Daten exportiert werden sollen des Assistenten für den Datenbank-Export die Option Neue Felder zu einer bestehenden Tabelle hinzufügen aus.
- Wählen Sie im Feld Tabelle oder Ansicht wählen die Datenbanktabelle aus.
- Gleichen Sie im Feld Fälle mit Datensätzen abgleichen die Variablen, die die einzelnen Fälle eindeutig identifizieren, mit den entsprechenden Datenbankfeldnamen ab.
- Verschieben Sie die Variablen, die als neue Felder hinzugefügt werden sollen, mittels Ziehen und Ablegen in die Spalte Quelle der Werte.

Abbildung 3-31
Assistent für den Datenbank-Export, Hinzufügen neuer Felder zu einer bestehenden Tabelle.



Informationen zu Feldnamen und Datentypen finden Sie im Abschnitt zum Erstellen von Datenbankfeldern aus SPSS-Variablen unter Export in eine Datenbank auf S. 54.

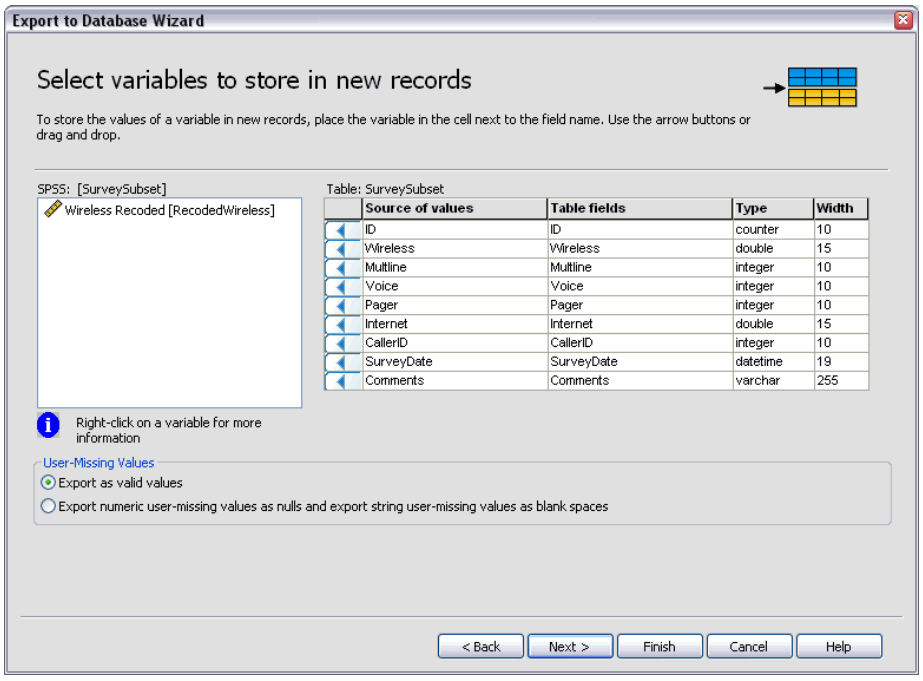
Bestehende Felder anzeigen. Wählen Sie diese Option, um eine Liste bestehender Felder anzuzeigen. Mit diesem Feld im Assistenten für den Datenbank-Export können Sie zwar keine bestehenden Felder ersetzen, sie können jedoch herausfinden, welche Felder bereits in der Tabelle vorhanden sind. Wenn Sie die Werte bestehender Felder ersetzen möchten, lesen Sie nach unter Ersetzen von Werten in bestehenden Feldern auf S. 61.

Anhängen neuer Datensätze (Fälle)

So können Sie neue Datensätze (Fälle) an eine Datenbanktabelle anhängen:

- Wählen Sie im Fenster Wählen Sie aus, wie die Daten exportiert werden sollen des Assistenten für den Datenbank-Export die Option Neue Datensätze an eine bestehende Tabelle anhängen aus.
- Wählen Sie im Feld Tabelle oder Ansicht wählen die Datenbanktabelle aus.
- Gleichen Sie die Variablen in der Arbeitsdatei mit den Tabellenfeldern ab, indem Sie die Variablen per Ziehen und Ablegen in die Spalte Quelle der Werte verschieben.

Abbildung 3-32
Assistent für den Datenbank-Export, Hinzufügen von Datensätzen (Fällen) zu einer Tabelle



Der Assistent für den Datenbank-Export wählt automatisch alle Variablen aus, die mit bestehenden Feldern übereinstimmen. Er verwendet dazu die in der Arbeitsdatei gespeicherten Informationen zu der ursprünglichen Datenbanktabelle (sofern verfügbar) und/oder Variablennamen, die mit Feldnamen identisch sind. Dieser erste automatische Abgleich ist nur als Richtlinie gedacht und hindert Sie nicht daran, die Art und Weise des Abgleichs der Variablen mit Datenbankfeldern zu ändern.

Beim Hinzufügen neuer Datensätze zu einer bestehenden Tabelle gelten folgende grundsätzlichen Regeln/Einschränkungen:

- Alle Fälle (bzw. alle ausgewählten Fälle) in der Arbeitsdatei werden zur Tabelle hinzugefügt. Wenn einer dieser Fälle mit einem bestehenden Datensatz in der Datenbank identisch ist, wird möglicherweise ein Fehler ausgegeben, sofern ein doppelter Schlüsselwert gefunden wird. Informationen zum Export ausgewählter Fälle finden Sie unter Auswahl der zu exportierenden Fälle auf S. 59.

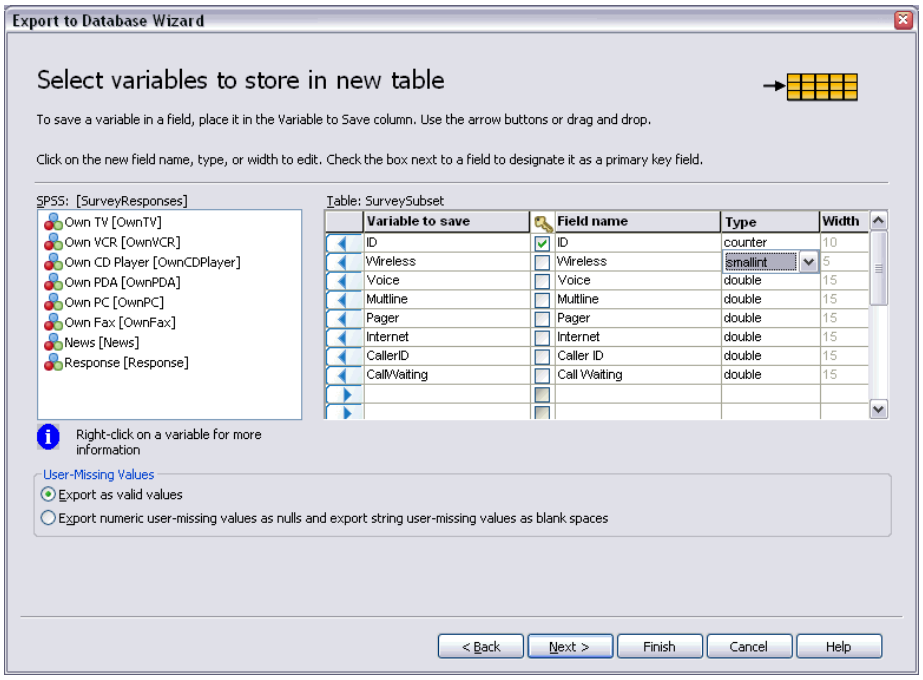
- Sie können die Werte neuer Variablen, die in dieser Sitzung erstellt wurden, als Werte für bestehende Felder verwenden. Sie können jedoch keine neuen Felder hinzufügen oder die Namen bestehender Felder ändern. Informationen zum Hinzufügen neuer Felder zu einer Datenbanktabelle finden Sie unter Hinzufügen neuer Felder auf S. 62.
- Alle ausgeschlossenen Datenbankfelder oder Felder, die keiner Variablen zugeordnet sind, weisen für die hinzugefügten Datensätze in der Datenbanktabelle keine Werte auf. (Wenn die Zelle Quelle der Werte leer ist, ist dem Feld keine Variable zugeordnet.)

Erstellen einer neuen Tabelle bzw. Ersetzen einer Tabelle

So erstellen Sie eine neue Datenbanktabelle bzw. ersetzen eine bestehende Datenbanktabelle:

- Wählen Sie im Fenster Wählen Sie aus, wie die Daten exportiert werden sollen des Exportassistenten die Option Eine bestehende Tabelle verwerfen und eine andere Tabelle mit demselben Namen erstellen aus oder wählen Sie die Option Neue Tabelle erstellen aus und geben Sie einen Namen für die neue Tabelle ein.
- Wenn Sie eine bestehende Tabelle ersetzen möchten, wählen Sie im Feld Tabelle oder Ansicht wählen die betreffende Datenbanktabelle aus.
- Verschieben Sie Variablen mithilfe von Ziehen und Ablegen in die Spalte Zu speichernde Variable.
- Wahlweise können Sie Variablen/Felder angeben, die den Primärschlüssel festlegen, Feldnamen ändern sowie den Datentyp ändern.

Abbildung 3-33
Assistent für den Datenbank-Export, Auswahl von Variablen für eine neue Tabelle



Primärschlüssel. Um Variablen als Primärschlüssel in der Datenbanktabelle anzugeben, aktivieren Sie das Kontrollkästchen in der mit dem Schlüsselsymbol gekennzeichneten Spalte.

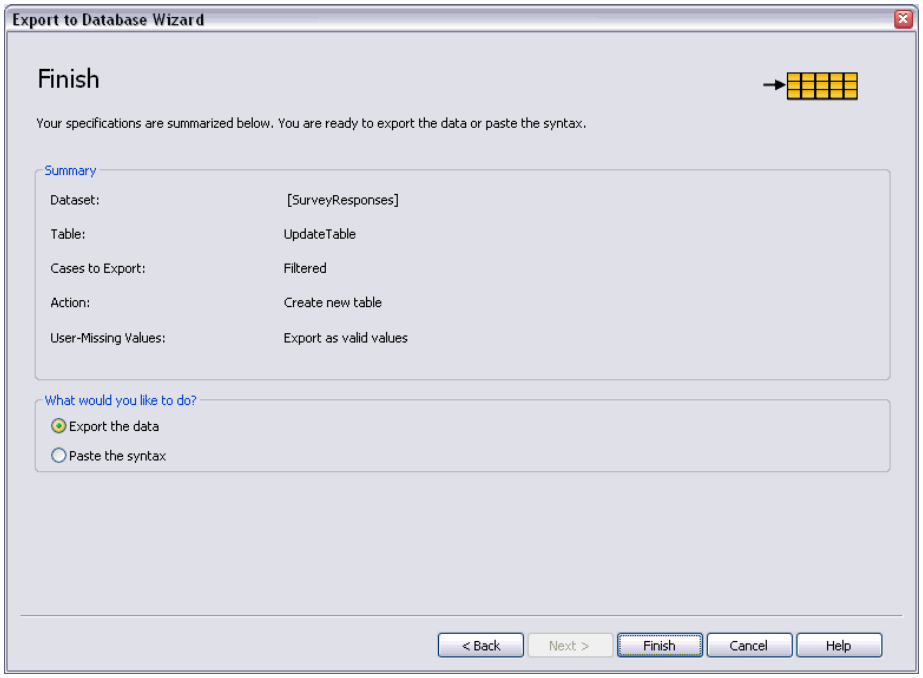
- Wenn nicht alle Werte des Primärschlüssels eindeutig sind, wird ein Fehler ausgegeben.
- Wenn Sie eine einzelne Variable als Primärschlüssel auswählen, muss jeder Datensatz (Fall) einen eindeutigen Wert für diese Variable aufweisen.
- Wenn Sie mehrere Variablen als Primärschlüssel auswählen, wird damit ein zusammengesetzter Primärschlüssel definiert und die *Kombination* der Werte für die ausgewählten Variablen muss für jeden Fall eindeutig sein.

Informationen zu Feldnamen und Datentypen finden Sie im Abschnitt zum Erstellen von Datenbankfeldern aus SPSS-Variablen unter Export in eine Datenbank auf S. 54.

Abschließen des Assistenten für den Datenbank-Export

Im letzten Fenster des Assistenten für den Datenbank-Export finden Sie eine Zusammenfassung, die angibt, welche Daten exportiert werden und wie der Exportvorgang erfolgen soll. Außerdem haben Sie die Möglichkeit, die Daten zu exportieren oder die zugrunde liegende Befehlssyntax in ein Syntaxfenster einzufügen.

Abbildung 3-34
Assistent für den Datenbank-Export, Abschlussfenster



Zusammenfassende Informationen

- **Daten-Set.** Der SPSS-Sitzungsname für das Daten-Set, das zum Exportieren der Daten verwendet werden soll. Diese Informationen sind vor allem dann nützlich, wenn Sie mehrere Datenquellen geöffnet haben. Über die grafische Benutzeroberfläche (z. B. den Datenbank-Assistenten) geöffneten Datenquellen werden automatisch Namen wie *DatenSet1*,

DatenSet2 usw. zugewiesen. Über Befehlssyntax geöffnete Datensätze tragen nur dann einen Daten-Set-Namen, wenn ein solcher Name explizit zugewiesen wurde.

- **Tabelle.** Der Name der zu bearbeitenden bzw. erstellenden Tabelle.
- **Zu exportierende Fälle.** Es können entweder alle Fälle exportiert werden oder nur die anhand einer zuvor definierten Filterbedingung ausgewählten Fälle. Für weitere Informationen siehe Auswahl der zu exportierenden Fälle auf S. 59.
- **Aktion.** Gibt an, wie die Datenbank geändert wird (z. B. Erstellen einer neuen Tabelle, Hinzufügen von Feldern oder Datensätzen zu einer bestehenden Tabelle).
- **Benutzerdefinierte fehlende Werte.** Benutzerdefiniert fehlende Werte können als gültige Werte exportiert oder (im Fall von numerischen Variablen) ebenso wie systemdefinierte Werte behandelt werden bzw. (im Fall von String-Variablen) in Leerzeichen konvertiert werden. Diese Einstellung wird in dem Feld festgelegt, in dem auch die zu exportierenden Variablen ausgewählt werden.

Export nach Dimensions

Im Dialogfeld "Export to Dimensions" wird eine SPSS-Datendatei und eine Dimensions-Metadatendatei erstellt, mit der Sie die Daten in Dimensions-Anwendungen, wie mrInterview und mrTables, einlesen können. Dies ist insbesondere dann hilfreich, wenn Sie Daten häufiger zwischen SPSS und Dimensions-Anwendungen austauschen. Sie können beispielsweise eine mrInterview-Datenquelle in SPSS einlesen, einige neue Variablen berechnen und dann die Daten in einem Format speichern, das von mrTables gelesen werden kann, ohne dass die ursprünglichen Metadatenattribute dabei verloren gehen.

So exportieren Sie Daten für die Verwendung in Dimensions-Anwendungen:

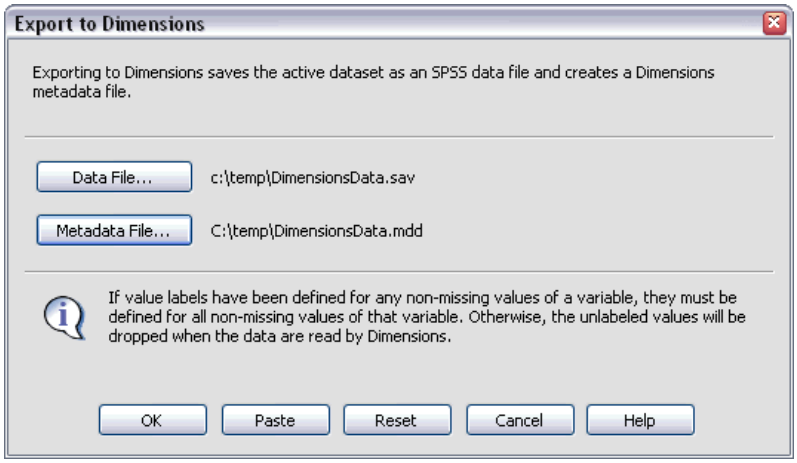
▶ Wählen Sie aus den Menüs in dem Fenster des Daten-Editors, das die zu exportierenden Daten enthält, folgende Optionen aus:

Datei
  Export nach Dimensions

▶ Klicken Sie auf Datendatei, um den Namen und den Standort der SPSS-Datendatei anzugeben.

▶ Klicken Sie auf Metadatendatei, um den Namen und den Standort der Dimensions-Metadatendatei anzugeben.

Abbildung 3-35
Dialogfeld "Export to Dimensions"



Für neue Variablen und Datensätze, die nicht aus Dimensions-Datenquellen erstellt wurden, werden SPSS-Variablenattribute Dimensions-Metadatenattributen in der Metadatendatei zugeordnet. Diese Zuordnung erfolgt nach den in der SPSS SAV DSC-Dokumentation in der Dimensions Development Library beschriebenen Methoden.

Wenn die Arbeitsdatei aus einer Dimensions-Datenquelle erstellt wurde:

- Die neue Metadatendatei wird durch Zusammenführung der ursprünglichen Metadatenattribute mit Metadatenattributen für etwaige neue Variablen erstellt, zuzüglich etwaiger Änderungen an ursprünglichen Variablen, die möglicherweise deren Metadatenattribute beeinflussen (z. B. Hinzufügen von oder Änderungen an Wertelabels).
- Bei Originalvariablen, die aus der Dimensions-Datenquelle eingelesen werden, bleiben alle Metadatenattribute, die von SPSS nicht erkannt werden, in ihrem ursprünglichen Zustand erhalten. So konvertiert SPSS beispielsweise Gittervariablen in reguläre SPSS-Variablen, die Metadaten, die diese Gittervariablen definieren, bleiben jedoch beim Speichern der neuen Metadatendatei erhalten.
- Wenn Dimensions-Variablen automatisch umbenannt wurden, damit Sie den Regeln für SPSS-Variablennamen entsprechen, ordnet die Metadatendatei die konvertierten Namen wieder den ursprünglichen Dimensions-Variablennamen zu.

Das Vorliegen bzw. Fehlen von Wertelabels kann die Metadatenattribute von Variablen beeinflussen und damit auch die Art und Weise, wie diese Variablen von den Dimensions-Anwendungen gelesen werden. Wenn für irgendwelche nichtfehlende Werte einer Variablen Wertelabels definiert wurden, müssen sie für alle nichtfehlenden Werte der betreffenden Variablen definiert werden. Anderenfalls werden die Werte ohne Label verworfen, wenn die Datei von Dimensions gelesen wird.

Schützen der ursprünglichen Daten

Um eine versehentliche Änderung oder Löschung der ursprünglichen Daten zu verhindern, können Sie die Datei mit einem Schreibschutz versehen.

▶ Wählen Sie die folgenden Menübefehle des Daten-Editors aus:

Datei
Datei als schreibgeschützt markieren

Wenn Sie danach Änderungen an den Daten vornehmen und versuchen, die Datendatei zu speichern, können Sie die Daten nur unter einem anderen Dateinamen speichern, sodass die ursprünglichen Daten unverändert erhalten bleiben.

Sie können die Dateiberechtigungen wieder auf "Lesen/Schreiben" zurücksetzen, indem Sie im Menü "Datei" die Option Datei für Lese-/Schreibzugriff markieren auswählen.

Virtuelle aktive Datei

Eine virtuelle aktive Datei ermöglicht die Arbeit mit großen Datendateien, ohne dass dabei mindestens so viel temporärer Speicherplatz auf der Festplatte vorhanden sein muss, wie die Größe der Datendatei beträgt. Bei den meisten Analyse- und Diagrammprozeduren wird die ursprüngliche Datenquelle bei jedem Ausführen erneut eingelesen. Bei Prozeduren, bei denen die Daten modifiziert werden, wird eine gewisse Menge an temporärem Speicherplatz auf der Festplatte benötigt, um die Änderungen aufzuzeichnen, und bei einigen Vorgängen ist immer mindestens ausreichend Speicherplatz auf der Festplatte für eine vollständige Kopie der Datendatei erforderlich.

Abbildung 3-36
Anforderungen an temporären Speicherplatz auf der Festplatte

Vorgang	GET FILE = 'v1-5 sav' FREQUENCIES...	COMPUTE v6 = ... RECODE v4... REGRESSION... /SAVE ZPRED	SORT CASES BY... oder CACHE
Virtuelle aktive Datei	v1 v2 v3 v4 v5 11 12 13 14 15 21 22 23 24 25 31 32 33 34 35 41 42 43 44 45 51 52 53 54 55 61 62 63 64 65	v1 v2 v3 v4 v5 v6 zpre 11 12 13 14 15 16 1 21 22 23 24 25 26 2 31 32 33 34 35 36 3 41 42 43 44 45 46 4 51 52 53 54 55 56 5 61 62 63 64 65 66 6	v1 v2 v3 v4 v5 v6 zpre 11 12 13 14 15 16 1 21 22 23 24 25 26 2 31 32 33 34 35 36 3 41 42 43 44 45 46 4 51 52 53 54 55 56 5 61 62 63 64 65 66 6
Vorübergehend auf der Festplatte gespeicherte Daten	Keine	v4 v6 zpre 14 16 1 24 26 2 34 36 3 44 46 4 54 56 5 64 66 6	v1 v2 v3 v4 v5 v6 zpre 11 12 13 14 15 16 1 21 22 23 24 25 26 2 31 32 33 34 35 36 3 41 42 43 44 45 46 4 51 52 53 54 55 56 5 61 62 63 64 65 66 6

Vorgänge, bei denen kein temporärer Speicherplatz auf der Festplatte benötigt wird:

- Einlesen von SPSS-Datendateien
- Zusammenfügen von zwei oder mehr SPSS-Datendateien
- Einlesen von Datenbankdateien mit dem Datenbank-Assistenten
- Zusammenfügen einer SPSS-Datendatei mit einer Datenbanktabelle
- Ausführen von Prozeduren, die Daten einlesen (beispielsweise Häufigkeiten, Kreuztabellen und Explorative Datenanalyse)

Vorgänge, bei denen mindestens eine Datenspalte in temporärem Speicherplatz auf der Festplatte erstellt wird:

- Berechnen von neuen Variablen
- Umkodieren von vorhandenen Variablen
- Ausführen von Prozeduren, bei denen Variablen erstellt oder modifiziert werden (beispielsweise das Speichern von vorhergesagten Werten bei der linearen Regression)

Vorgänge, bei denen eine vollständige Kopie der Datendatei in temporärem Speicherplatz auf der Festplatte erstellt wird:

- Einlesen von Excel-Dateien
- Ausführen von Prozeduren zum Sortieren von Daten (beispielsweise die Prozeduren "Fälle sortieren" und "Datei aufteilen")
- Einlesen von Daten mit den Syntaxbefehlen `GET TRANSLATE` und `DATA LIST`
- Verwenden der Funktionen für das Ablegen von Daten im Zwischenspeicher oder des Syntaxbefehls `CACHE`
- Starten von anderen Anwendungen aus SPSS heraus, die die Datendatei einlesen (beispielsweise AnswerTree und DecisionTime)

Anmerkung: Der Syntaxbefehl `GET DATA` stellt ähnliche Funktionen wie der Syntaxbefehl `DATA LIST` bereit, erstellt jedoch keine vollständige Kopie der Datendatei in temporärem Speicherplatz auf der Festplatte. Mit dem Syntaxbefehl `SPLIT FILE` werden die Daten in der Datendatei nicht sortiert. Deshalb wird auch keine Kopie der Datendatei erstellt. Damit dieser Befehl jedoch ordnungsgemäß ausgeführt werden kann, müssen die Daten sortiert sein. Über die Benutzeroberfläche des Dialogfelds dieser Prozedur wird die Datendatei automatisch sortiert und eine vollständige Kopie der Datendatei erstellt. (Die Befehlssyntax ist in der Studentenversion nicht verfügbar.)

Vorgänge, bei denen in der Standardeinstellung eine vollständige Kopie der Datendatei erstellt wird:

- Einlesen von Datenbanken mit dem Datenbank-Assistenten
- Einlesen von Textdateien mit dem Text-Assistenten

Der Text-Assistent bietet eine optionale Einstellung zum automatischen Zwischenspeichern der Daten. Diese Option ist standardmäßig ausgewählt. Sie können diese Auswahl aufheben, indem Sie das Kontrollkästchen Daten in lokalen Zwischenspeicher deaktivieren. Beim Datenbank-Assistenten können Sie die erstellte Befehlssyntax einfügen und den Befehl `CACHE` löschen.

Erstellen eines Zwischenspeichers für Daten

Die virtuelle aktive Datei kann die benötigte Menge an temporärem Speicherplatz auf der Festplatte drastisch reduzieren. Das Nichtvorhandensein einer temporären Kopie der eigentlich aktiven Datei bedeutet aber auch, dass die ursprüngliche Datendatei für jede Prozedur neu eingelesen werden muss. Bei großen Datendateien, die aus einer externen Quelle eingelesen werden, kann das Erstellen einer temporären Kopie der Daten die Leistung steigern. Bei Tabellen in einer Datenbank beispielsweise muss die SQL-Abfrage, mit der die Informationen aus der

Datenbank ausgelesen werden, für jeden Befehl und jede Prozedur erneut ausgeführt werden, bei denen Daten eingelesen werden. Da fast alle Statistik- und Diagrammprozeduren die Daten einlesen müssen, wird die SQL-Abfrage für jede aufgerufene Prozedur erneut ausgeführt. Bei einer großen Anzahl an Prozeduren kann dies zu einer beträchtlichen Steigerung der für die Verarbeitung benötigten Zeit führen.

Wenn auf dem Computer, auf dem die Analyse durchgeführt wird (der lokale Computer oder der Remote-Server), ausreichend Speicherplatz auf der Festplatte vorhanden ist, können Sie die mehrfache Ausführung von SQL-Afragen vermeiden und somit die Verarbeitungszeit verringern, indem Sie einen Zwischenspeicher für die Daten aus der aktiven Datei anlegen. Der Zwischenspeicher für die Daten ist eine temporäre Kopie der gesamten Daten.

Anmerkung: In der Standardeinstellung erstellt der Datenbank-Assistent automatisch einen Zwischenspeicher für die Daten. Wenn Sie aber mithilfe des Syntaxbefehls `GET DATA` eine Datenbank einlesen, wird nicht automatisch ein Zwischenspeicher für die Daten erstellt. (Die Befehlssyntax ist in der Studentenversion nicht verfügbar.)

So erstellen Sie einen Zwischenspeicher für die Daten:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Daten in Zwischenspeicher...

▶ Klicken Sie auf OK oder Jetzt zwischenspeichern.

Mit OK werden die Daten in den Zwischenspeicher überführt, wenn das Programm die Daten das nächste Mal einliest, beispielsweise beim nächsten Ausführen einer statistischen Prozedur. In der Regel empfiehlt sich dieses Vorgehen, da hierbei kein zusätzlicher Aufwand beim Einlesen der Daten entsteht. Mit Jetzt zwischenspeichern werden die Daten sofort in den Zwischenspeicher übertragen. In den meisten Situationen ist dies nicht notwendig. Jetzt zwischenspeichern ist primär aus zwei Gründen nützlich:

- Eine Datenquelle ist gesperrt und kann nicht durch andere aktualisiert werden, bis Sie Ihre Sitzung beenden, eine andere Datenquelle öffnen oder die Daten zwischenspeichern.
- Bei umfangreichen Datenquellen erfolgt der Bildlauf durch den Inhalt der Registerkarte "Datenansicht" des Daten-Editors viel schneller, wenn Sie die Daten zwischenspeichern.

So können Sie Daten automatisch zwischenspeichern:

Mithilfe des Befehls `SET` können Sie nach einer festgelegten Anzahl von Änderungen in der aktiven Datendatei automatisch einen Zwischenspeicher für die Daten erstellen. In der Standardeinstellung wird die aktive Datendatei nach 20 Änderungen automatisch zwischengespeichert.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Neu
    Syntax

▶ Geben Sie im Syntax-Fenster Folgendes ein: SET CACHE n. (Dabei steht *n* für die Anzahl der Änderungen in der aktiven Datendatei, nach der die Datendatei zwischengespeichert wird.)

- Wählen Sie die folgenden Befehle aus den Menüs des Syntax-Fensters aus:

  Ausführen
  Alles

  Anmerkung: Die Einstellung für den Zwischenspeicher wird nicht für alle Sitzungen übernommen. Bei jedem Start einer neuen Sitzung wird der Wert auf den Standardwert 20 zurückgesetzt.

Kapitel
4

Modus für verteilte Analysen

Beim Modus für verteilte Analysen können Sie speicherintensive Vorgänge von Ihrem lokalen Computer weg auf einen anderen Computer auslagern. Da die für verteilte Analysen eingesetzten Remote-Server in der Regel leistungsfähiger und schneller als Ihr lokaler Computer sind, kann ein sinnvoller Einsatz des Modus für verteilte Analysen die für die Verarbeitung benötigte Zeit beträchtlich verringern. In den folgenden Situationen kann die verteilte Analyse auf einem Remote-Server nützlich sein:

- Sie arbeiten mit umfangreichen Datendateien oder mit Daten, die aus einer Datenbank eingelesen werden.
- Für die Analyse müssen speicherintensive Aufgaben durchgeführt werden. Alle Aufgaben, deren Verarbeitung im Modus für lokale Analysen sehr lange dauert, können möglicherweise von der verteilten Analyse profitieren.

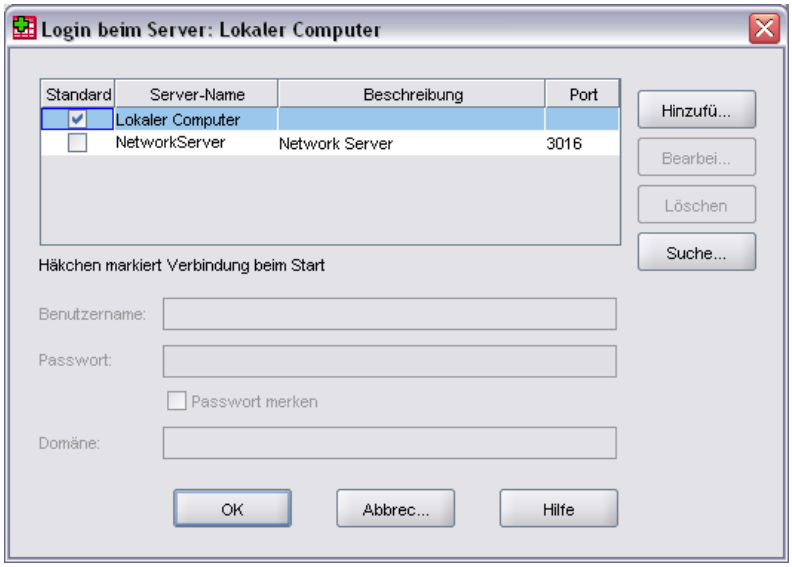
Die verteilte Analyse greift nur bei datenbezogenen Aufgaben. Hierzu gehören beispielsweise das Einlesen von Daten, das Transformieren von Daten, das Berechnen neuer Variablen und das Berechnen von Statistiken. Die verteilte Analyse hat keine Auswirkungen auf Aufgaben im Zusammenhang mit der Bearbeitung der Ausgaben. Hierzu gehören beispielsweise das Bearbeiten von Pivot-Tabellen oder das Modifizieren von Diagrammen.

Anmerkung: Der Modus für verteilte Analysen steht nur zur Verfügung, wenn sowohl eine lokale Version von SPSS als auch der Zugriff auf eine lizenzierte Server-Version von SPSS vorliegt, die auf einem Remote-Server installiert ist. Außerdem stehen verteilte Analysen nur für Windows-Desktop-Computer zur Verfügung. Für Mac und Linux sind sie nicht verfügbar.

Login beim SPSS-Server

Im Dialogfeld “Login beim Server” können Sie auswählen, welcher Computer Befehle verarbeiten und Prozeduren ausführen soll. Sie können Ihren lokalen Computer oder einen Remote-Server auswählen.

Abbildung 4-1
Dialogfeld "Login beim SPSS-Server"



Sie können der Liste Remote-Server hinzufügen, Remote-Server in der Liste ändern und aus dieser entfernen. Für Remote-Server ist in der Regel ein Benutzername und ein Passwort erforderlich. Möglicherweise müssen Sie außerdem einen Domänennamen angeben. Wenden Sie sich an Ihren Systemadministrator, um Informationen zu verfügbaren Servern, Benutzernamen, Paßwörtern, Domänennamen und andere Verbindungsinformationen zu erhalten.

Sie können einen Standard-Server auswählen und die Benutzernamen, Domänennamen und Passwörter für beliebige Server speichern. Beim Starten einer neuen Sitzung wird automatisch eine Verbindung mit dem Standard-Server hergestellt.

Wenn Sie über eine Lizenz zur Verwendung von SPSS Adaptor für Predictive Enterprise Services verfügen und an Ihrem Standort SPSS Predictive Enterprise Services 3.5 oder höher ausgeführt wird, können Sie durch Klicken auf Suche... eine Liste der Server anzeigen, die in Ihrem Netzwerk verfügbar sind. Wenn Sie nicht bei einem Predictive Enterprise-Repository angemeldet sind, werden Sie zur Eingabe von Verbindungsinformationen aufgefordert, bevor die Liste der Server angezeigt wird.

Hinzufügen und Bearbeiten von Einstellungen für die Server-Anmeldung

Verwenden Sie das Dialogfeld "Einstellungen für Server-Anmeldung", um Verbindungsinformationen für Remote-Server im Modus für verteilte Analysen hinzuzufügen bzw. zu bearbeiten.

Abbildung 4-2
Dialogfeld "Einstellungen für Server-Anmeldung"

Wenden Sie sich an Ihren Systemadministrator, um Informationen zu den verfügbaren Servern, Portnummern für diese Server und weitere Verbindungsinformationen zu erhalten. Verwenden Sie Secure Socket Layer nur, wenn Sie von Ihrem Administrator dazu angewiesen wurden.

Server-Name. Der Name des Servers kann ein dem Computer zugewiesener alphanumerischer Name (beispielsweise "NetzwerkServer") oder eine dem Computer zugewiesene eindeutige IP-Adresse sein (beispielsweise 202.123.456.78).

Portnummer. Die Portnummer bezeichnet den Port, den die Serversoftware für die Kommunikation verwendet.

Beschreibung. Sie können eine optionale Beschreibung eingeben, die in der Serverliste angezeigt werden soll.

Stellen Sie eine Verbindung mit Secure Socket Layer her. Secure Socket Layer (SSL) verschlüsselt Anforderungen für verteilte Analysen, wenn diese an den SPSS-Remote-Server gesendet werden. Verwenden Sie SSL nicht, ohne zuvor mit Ihrem Administrator Rücksprache gehalten zu haben. SSL muss auf Ihrem Desktop-Computer und auf dem Server konfiguriert sein, damit diese Option aktiviert werden kann.

So wählen Sie einen Server aus, wechseln den Server oder fügen einen neuen Server hinzu:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Server umschalten...

So wählen Sie einen Standardserver aus:

▶ Wählen Sie in der Serverliste das Kästchen neben dem Server aus, den Sie verwenden möchten.

▶ Geben Sie den Benutzernamen, Domänennamen und das Passwort ein, die Sie vom Administrator erhalten haben.

Anmerkung: Beim Starten einer neuen Sitzung wird automatisch eine Verbindung mit dem Standard-Server hergestellt.

So schalten Sie auf einen anderen Server um:

- ▶ Wählen Sie einen Server aus der Liste aus.
- ▶ Geben Sie erforderlichenfalls Ihren Benutzernamen, den Domänennamen und Ihr Paßwort ein.

Anmerkung: Wenn Sie in einer Sitzung den Server wechseln, werden alle geöffneten Fenster geschlossen. Ehe die Fenster geschlossen werden, werden Sie aufgefordert, die vorgenommenen Änderungen zu speichern.

So fügen Sie einen Server hinzu:

- ▶ Besorgen Sie sich die Verbindungsinformationen für den Server vom Administrator.
- ▶ Klicken Sie auf Hinzufügen, um das Dialogfeld "Einstellungen für Server-Anmeldung" zu öffnen.
- ▶ Geben Sie die Verbindungsinformationen und optionalen Einstellungen ein und klicken Sie anschließend auf OK.

So bearbeiten Sie einen Server:

- ▶ Wenden Sie sich an den Administrator, um die geänderten Verbindungsinformationen für den Server zu erhalten.
- ▶ Klicken Sie auf Bearbeiten, um das Dialogfeld "Einstellungen für Server-Anmeldung" zu öffnen.
- ▶ Geben Sie die Änderungen ein und klicken Sie anschließend auf OK.

So können Sie nach den verfügbaren Servern suchen:

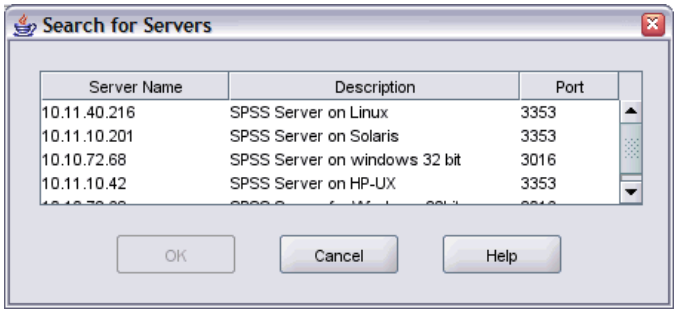
Anmerkung: Die Möglichkeit nach verfügbaren Servern zu suchen, steht nur zur Verfügung, wenn Sie über eine Lizenz zur Verwendung von SPSS Adaptor für Predictive Enterprise Services verfügen und an Ihrem Standort SPSS Predictive Enterprise Services 3.5 oder höher ausgeführt wird.

- ▶ Klicken Sie auf Suche..., um das Dialogfeld "Search for Servers" zu öffnen. Wenn Sie nicht bei einem Predictive Enterprise-Repository angemeldet sind, werden Sie zur Eingabe von Verbindungsinformationen aufgefordert.
- ▶ Wählen Sie einen oder mehrere verfügbare Server aus und klicken Sie auf OK. Die Server werden nun im Dialogfeld "Login beim Server" angezeigt.
- ▶ Um eine Verbindung mit einem der Server herzustellen, befolgen Sie die Anweisungen unter "So schalten Sie auf einen anderen Server um".

Suche nach verfügbaren Servern

Verwenden Sie das Dialogfeld "Search for Servers", um einen oder mehrere Server auszuwählen, die in Ihrem Netzwerk verfügbar sind. Dieses Dialogfeld wird angezeigt, wenn Sie im Dialogfeld "Login beim Server" auf Suche... klicken.

Abbildung 4-3
Dialogfeld "Search for Servers"



Wählen Sie einen oder mehrere Server aus und klicken Sie auf OK, um sie zum Dialogfeld "Login beim Server" hinzuzufügen. Sie können zwar auch manuell Server im Dialogfeld "Login beim Server" hinzufügen, doch die Suche nach verfügbaren Servern ermöglicht Ihnen, eine Verbindung zu Servern herzustellen, ohne dass Ihnen der richtige Servername und die Portnummer bekannt sein muss. Diese Informationen werden automatisch bereitgestellt. Sie benötigen jedoch auch in diesem Fall die richtigen Anmeldeinformationen, wie Benutzername, Domäne und Passwort.

Öffnen von Datendateien auf einem Remote-Server

Im Modus für verteilte Analysen wird statt des Standard-Dialogfelds "Datei öffnen" das Dialogfeld "Entfernte Datei öffnen" angezeigt.

- Die Inhalte in der Liste der verfügbaren Dateien, Ordner und Laufwerke hängen davon ab, was auf dem Remote-Server bzw. vom Remote-Server aus verfügbar ist. Der Name des aktuellen Servers wird im oberen Teil des Dialogfelds angezeigt.
- Im Modus für verteilte Analysen können Sie nur auf Dateien auf dem lokalen Computer zugreifen, wenn Sie das Laufwerk oder die Ordner mit den Datendateien für den gemeinsamen Zugriff freigeben. Informationen zur Freigabe von Ordnern auf ihrem lokalen Computer für den gemeinsamen Zugriff im Servernetzwerk finden Sie in der Dokumentation zu Ihrem Betriebssystem.
- Wird auf dem Server ein anderes Betriebssystem ausgeführt (Ihr Computer läuft beispielsweise unter Windows und der Server unter UNIX), werden Sie im Modus für verteilte Analysen wahrscheinlich keinen Zugriff auf lokale Datendateien haben, selbst wenn sie sich in freigegebenen Ordnern befinden.

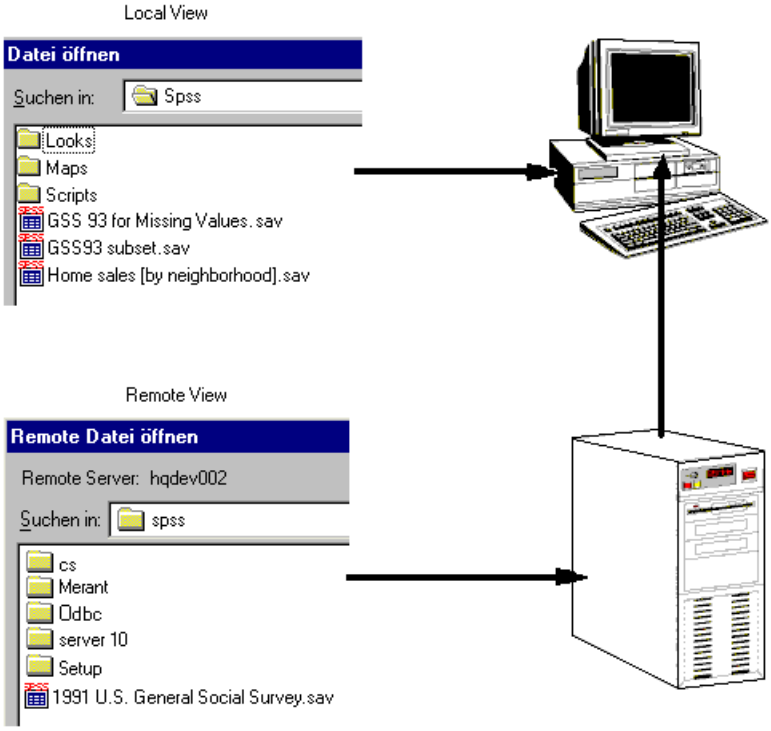
Dateizugriff im Modus für lokale und verteilte Analysen

Es hängt von dem Computer ab, den Sie zum Verarbeiten von Befehlen und Ausführen von Prozeduren verwenden, welche Datenordner (Verzeichnisse) und Laufwerke sowohl auf dem lokalen Computer als auch im Netzwerk angezeigt werden. Denken Sie daran, dass es sich bei diesem Computer nicht notwendigerweise um den Computer handelt, an dem Sie arbeiten.

Modus für lokale Analysen. Wenn Sie Ihren lokalen Computer als “Server” einsetzen, werden im Dialogfeld zum Öffnen von Dateien dieselben Datendateien, Ordner und Laufwerke angezeigt wie in anderen Anwendungen oder im Windows-Explorer. Es werden alle Datendateien und Ordner auf Ihrem Computer sowie alle Dateien und Ordner auf verbundenen Netzlaufwerken angezeigt.

Modus für verteilte Analysen. Wenn Sie einen anderen Computer als Remote-Server zum Verarbeiten von Befehlen und Ausführen von Prozeduren einsetzen, werden Datendateien, Ordner und Laufwerke aus der Sicht des Remote-Servers angezeigt. Möglicherweise werden bekannte Ordnernamen wie *Programme* und Laufwerke wie *C:* angezeigt. Hierbei handelt es sich aber *nicht* um Ordner und Laufwerke auf dem lokalen Computer, sondern um Ordner und Dateien auf dem Remote-Server.

Abbildung 4-4
Lokale und Remote-Ansichten



Im Modus für verteilte Analysen können Sie nur auf Datendateien auf dem lokalen Computer zugreifen, wenn Sie das Laufwerk oder die Ordner mit den Datendateien für den gemeinsamen Zugriff freigeben. Wird auf dem Server ein anderes Betriebssystem ausgeführt (Ihr Computer läuft beispielsweise unter Windows und der Server unter UNIX), werden Sie im Modus für verteilte Analysen wahrscheinlich keinen Zugriff auf lokale Datendateien haben, selbst wenn sie sich in freigegebenen Ordnern befinden.

Der Modus für verteilte Analysen entspricht nicht dem einfachen Zugriff auf Datendateien, die sich auf einem anderen Computer im Netzwerk befinden. Der Zugriff auf Datendateien, die auf einem anderen Gerät im Netzwerk gespeichert sind, kann sowohl im Modus für lokale Analysen als auch im Modus für verteilte Analysen stattfinden. Im lokalen Modus können Sie von Ihrem

lokalen Computer aus auf andere Geräte im Netzwerk zugreifen. Im verteilten Modus können Sie vom Remote-Server aus auf andere Geräte im Netz zugreifen.

Wenn Sie nicht sicher sind, ob Sie im Modus für lokale Analysen oder im Modus für verteilte Analysen arbeiten, schauen Sie in der Titelleiste eines der Dialogfelder für den Zugriff auf die Daten nach. Wenn der Titel des Dialogfelds das Wort *Entfernt* (wie beispielsweise Entfernte Datei öffnen) enthält oder der Text Remote-Server: [Servername] im oberen Teil des Dialogfelds angezeigt wird, arbeiten Sie im Modus für verteilte Analysen.

Anmerkung: Dies gilt nur für Dialogfelder für den Zugriff auf Datendateien (beispielsweise zum Öffnen und Speichern von Daten, zum Öffnen von Datenbanken und zum Zuweisen des Datenlexikons). Bei allen anderen Dateitypen (beispielsweise Viewer-Dateien, Syntaxdateien und Skriptdateien) werden jeweils die lokal gespeicherten Dateien gezeigt.

Verfügbarkeit von Prozeduren im Modus für verteilte Analysen

Im Modus für verteilte Analysen können Prozeduren nur verwendet werden, wenn diese sowohl auf dem lokalen Computer als auch auf dem Remote-Server installiert sind.

Wenn Sie optionale Komponenten lokal installiert haben, die auf dem Remote-Server nicht zur Verfügung stehen, und Sie zwischen dem lokalen Computer und dem Remote-Server wechseln, werden die entsprechenden Prozeduren aus dem Menü entfernt und die Befehlssyntax wird lediglich zu Fehlern führen. Durch einen Wechsel zurück in den lokalen Modus werden die betroffenen Prozeduren wiederhergestellt.

Absolute und relative Pfadangaben

Im Modus für verteilte Analysen sind relative Pfadangeben für Datendateien und Befehlssyntaxdateien relativ zum aktuellen Server, nicht relativ zum lokalen Computer. Eine relative Pfadangabe (wie */mydocs/mydata.sav*) verweist nicht auf ein Verzeichnis und eine Datei auf Ihrem lokalen Laufwerk sondern auf ein Verzeichnis und eine Datei auf der Festplatte des Remote-Servers.

Windows-UNC-Pfadangaben

Wenn Sie eine Version für Windows-Server verwenden, können Sie beim Zugriff auf Datendateien und Syntaxdateien mit der Befehlssyntax UNC-Pfadangaben (UNC = Universal Naming Convention, Universelle Namenskonvention) verwenden. UNC-Pfadangaben weisen die folgende allgemeine Form auf:

```
\\Servername\Freigabe\Pfad\Dateiname
```

- *Servername* ist der Name des Computers, auf dem die Datendatei gespeichert ist.
- *Freigabe* ist der Ordner (das Verzeichnis) auf diesem Computer, der (oder das) freigegeben ist.
- *Pfad* sind die dem freigegebenen Verzeichnis untergeordneten Ordner (Unterordner bzw. Unterverzeichnisse).
- *Dateiname*ist der Name der Datendatei.

Ein Beispiel lautet folgendermaßen:

```
GET FILE='\\hqdev001\public\july\sales.sav'.
```

Wenn dem Computer kein Name zugewiesen wurde, können Sie seine IP-Adresse verwenden, wie in folgendem Beispiel:

```
GET FILE='\\204.125.125.53\public\july\sales.sav'.
```

Auch mit UNC-Pfadangaben können Sie nur auf Datendateien und Syntaxdateien zugreifen, die sich auf Geräten oder in Ordnern befinden, die ausdrücklich freigegeben wurden. Beim Modus für verteilte Analysen gilt dies auch für die Datendateien und Syntaxdateien auf Ihrem lokalen Computer.

Absolute Pfadangaben unter UNIX

Auf Versionen für UNIX-Server gibt es kein Äquivalent zu UNC-Pfaden. Alle Verzeichnispfade müssen absolute Pfade sein, die beim Stamm des Servers beginnen; relative Pfade sind unzulässig. Wenn die Datendatei beispielsweise unter */bin/spss/data* gespeichert ist und das aktuelle Verzeichnis ebenfalls */bin/spss/data* ist, dann ist `GET FILE='sales.sav'` unzulässig; Sie müssen den gesamten Pfad angeben:

```
GET FILE='/bin/spss/sales.sav'.
INSERT FILE='/bin/spss/salesjob.sps'.
```

Kapitel
5

Daten-Editor

Der Daten-Editor bietet eine praktische Methode zum Erstellen und Bearbeiten von SPSS-Datendateien, die der von Tabellenkalkulationen bekannten Methode ähnelt. Das Fenster des Daten-Editors wird automatisch geöffnet, wenn Sie eine SPSS-Sitzung beginnen.

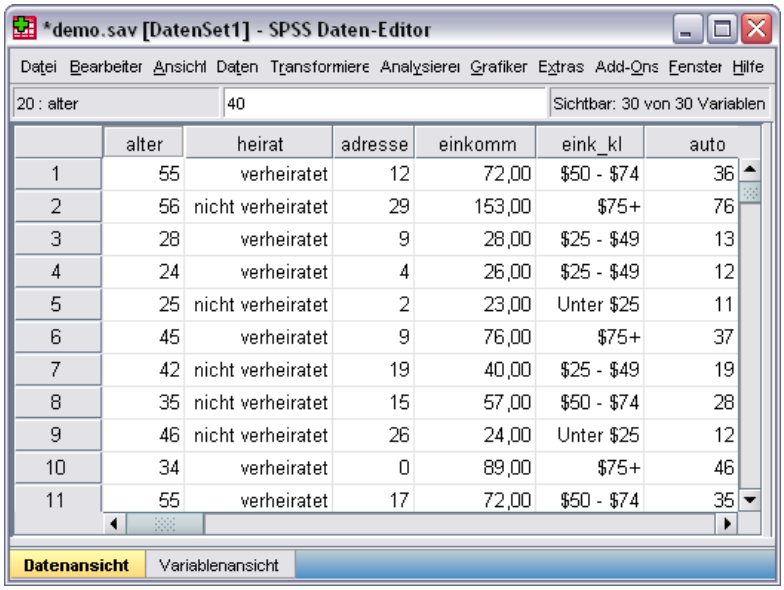
Der Daten-Editor stellt zwei Ansichten der Daten bereit:

- **Datenansicht.** In dieser Ansicht werden die eigentlichen Datenwerte oder die definierten Wertelabels angezeigt.
- **Variablenansicht.** In dieser Ansicht werden Informationen zu den Variablendefinitionen angezeigt. Dies umfasst die Variablen- und Wertelabels, den Datentyp (beispielsweise String, Datum oder numerisch), das Messniveau (nominal, ordinal oder metrisch) sowie benutzerdefinierte fehlende Werte.

In beiden Ansichten können Sie der Datendatei neue Informationen hinzufügen oder vorhandene Informationen bearbeiten und löschen.

Datenansicht

Abbildung 5-1
Datenansicht



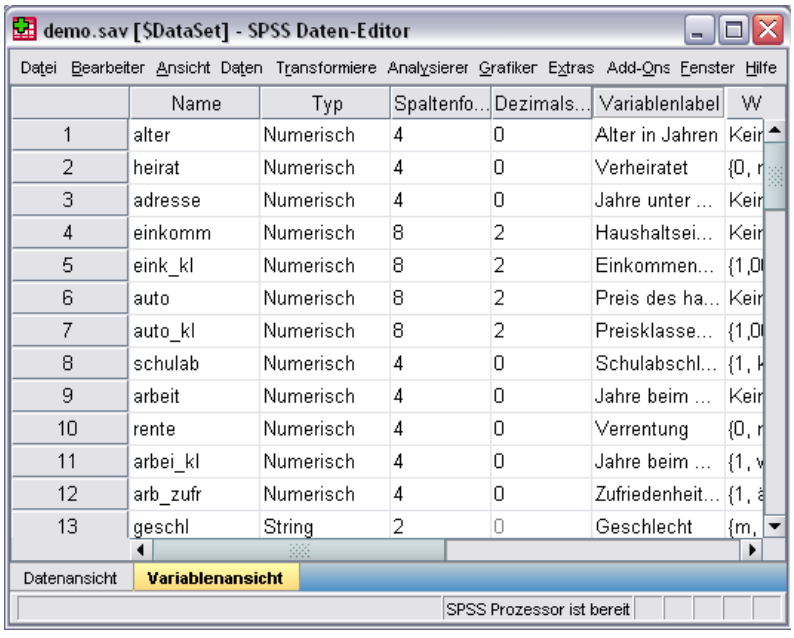
	alter	heirat	adresse	einkomm	eink_kl	auto
1	55	verheiratet	12	72,00	$50 - $74	36
2	56	nicht verheiratet	29	153,00	$75+	76
3	28	verheiratet	9	28,00	$25 - $49	13
4	24	verheiratet	4	26,00	$25 - $49	12
5	25	nicht verheiratet	2	23,00	Unter $25	11
6	45	verheiratet	9	76,00	$75+	37
7	42	nicht verheiratet	19	40,00	$25 - $49	19
8	35	nicht verheiratet	15	57,00	$50 - $74	28
9	46	nicht verheiratet	26	24,00	Unter $25	12
10	34	verheiratet	0	89,00	$75+	46
11	55	verheiratet	17	72,00	$50 - $74	35

Viele Funktionen der Datenansicht ähneln den Funktionen von Anwendungen für die Tabellenkalkulation. Es gibt allerdings mehrere wichtige Unterschiede:

- Zeilen sind Fälle. Jede Zeile stellt einen Fall oder eine Beobachtung dar. So ist zum Beispiel jede Person, die einen Fragebogen ausfüllt, ein Fall.
- Spalten sind Variablen. Jede Spalte stellt eine Variable oder eine Eigenschaft dar, die gemessen wurde. Jedes Objekt auf einem Fragebogen ist zum Beispiel eine Variable.
- Zellen enthalten Werte. Jede Zelle enthält einen einzelnen Wert einer Variablen für einen Fall. Die Zelle befindet sich an der Schnittstelle von Fall und Variable. Zellen enthalten nur Datenwerte. Im Gegensatz zu Programmen für die Tabellenkalkulation können Zellen im Daten-Editor keine Formeln enthalten.
- In einer Datendatei enthalten alle Zeilen die gleiche Anzahl Zellen. Die Dimensionen der Datendatei werden von der Anzahl der Fälle und Variablen bestimmt. In alle Zellen können Daten eingegeben werden. Wenn Sie Daten in eine Zelle außerhalb der Grenzen der definierten Datendatei eingeben, erweitert SPSS das Datenfeld, sodass es alle Zeilen und/oder Spalten einschließt, die zwischen dieser Zelle und den Grenzen der Datendatei liegen. Innerhalb der Grenzen der Datendatei gibt es keine “leeren” Zellen. Bei numerischen Variablen werden leere Zellen zum systemdefinierten fehlenden Wert konvertiert. Bei String-Variablen gelten leere Felder als gültiger Wert.

Variablenansicht

Abbildung 5-2
Variablenansicht



In der Variablenansicht werden die Attribute aller Variablen in der Datendatei angezeigt. In der Variablenansicht gilt Folgendes:

- Die Zeilen stellen Variablen dar.
- Die Spalten stellen die Attribute der Variablen dar.

Sie können Variablen hinzufügen und löschen, und Sie können die folgenden Variablenattribute ändern:

- Variablenname
- Datentyp
- Anzahl Ziffern oder Zeichen
- Anzahl Dezimalstellen
- Beschreibende Variablen- und Wertelabels
- Benutzerdefinierte fehlende Werte
- Spaltenbreite
- Messniveau

Alle diese Attribute werden beim Speichern der Datendatei gespeichert.

Neben den Methoden, mit denen Variableneigenschaften in der Variablenansicht definiert werden, gibt es zwei weitere Methoden zum Definieren von Variableneigenschaften:

- Mit dem Assistenten zum Kopieren von Dateneigenschaften können Sie eine externe SPSS-Datendatei oder ein anderes Daten-Set, das in der aktuellen Sitzung verfügbar ist, als Vorlage für die Definition von Datei- und Variableneigenschaften in der Arbeitsdatei verwenden. Sie können außerdem Variablen in der Arbeitsdatei als Vorlagen für andere Variablen in der Arbeitsdatei verwenden. Sie können den Assistenten zum Kopieren von Dateneigenschaften starten, indem Sie im Fenster des Daten-Editors im Menü “Daten” den Befehl “Dateneigenschaften kopieren” auswählen.
- Mit dem Befehl “Variableneigenschaften definieren” (ebenfalls im Menü “Daten” des Daten-Editors) können Sie Ihre Daten durchsuchen und eine Liste mit allen eindeutigen Datenwerten für die ausgewählten Variablen erstellen, Werte ohne Labels ausfindig machen und Werte automatisch mit Labels versehen. Diese Methode ist insbesondere für kategoriale Variablen sinnvoll, bei denen numerische Codes Kategorien darstellen, beispielsweise 0 = *männlich* und 1 = *weiblich*.

So zeigen Sie die Attribute von Variablen an und legen diese fest:

- ▶ Aktivieren Sie das Fenster des Daten-Editors.
- ▶ Doppelklicken Sie in der Datenansicht auf den Namen einer Variablen oben in einer Spalte oder klicken Sie auf die Registerkarte Variablenansicht.
- ▶ Wenn Sie eine neue Variable definieren möchten, geben Sie einen Namen in eine beliebige leere Zeile ein.
- ▶ Wählen Sie die Attribute aus, die Sie festlegen oder ändern möchten.

Variablennamen

Beim Benennen von Variablen gelten die folgenden Regeln:

- Variablennamen müssen eindeutig sein. Doppelt vorkommende Namen sind nicht zulässig.
- Variablennamen können bis zu 64 Byte lang sein. Das erste Zeichen muss ein Buchstabe oder eines der folgenden Zeichen sein: @, #, $. Bei den nachfolgenden Zeichen kann es sich um eine beliebige Kombination aus Buchstaben, Zahlen, einem Punkt (.) sowie um andere Zeichen handeln, die nicht in der Zeichensetzung verwendet werden. Im Codeseitenmodus entsprechen 64 Byte in Single-Byte-Sprachen (z. B. Englisch, Französisch, Deutsch, Spanisch, Italienisch, Hebräisch, Russisch, Griechisch, Arabisch, Thai) normalerweise 64 Zeichen und in Double-Byte-Sprachen (z. B. Japanisch, Chinesisch und Koreanisch) normalerweise 32 Zeichen. Viele String-Zeichen, die im Codeseitenmodus normalerweise nur ein Byte umfassen, umfassen im Unicode-Modus zwei oder mehr Byte. So umfasst é ein Byte im Codeseitenformat, aber zwei Byte im Unicode-Format; *résumé* ist also in einer Codeseitendatei sechs Byte, im Unicode-Modus dahingegen acht Byte lang.

  Anmerkung: Buchstaben umfassen alle Zeichen, die nicht der Zeichensetzung dienen und die zum Schreiben üblicher Wörter in den Sprachen verwendet werden, die von dem Zeichensatz der Plattform unterstützt werden.
- Variablennamen dürfen keine Leerzeichen enthalten.
- Das Zeichen # an der ersten Stelle eines Variablennamens definiert eine Arbeitsvariable. Arbeitsvariablen können nur mit Befehlssyntax erstellt werden. # kann nicht als erstes Zeichen einer Variablen in Dialogfeldern angegeben werden, die zum Erstellen neuer Variablen dienen.
- Ein $-Zeichen an der ersten Stelle zeigt an, dass es sich bei der Variablen um eine Systemvariable handelt. Das $-Zeichen ist nicht als erstes Zeichen von benutzerdefinierten Variablen zulässig.
- Punkt, Unterstrich und die Zeichen $, # und @ können in Variablennamen verwendet werden. So ist beispielsweise *A._$@#1* ein gültiger Variablenname.
- Variablennamen, die mit einem Punkt enden, sollten vermieden werden, da der Punkt als Befehlsabschluss interpretiert werden kann. Variablen, die mit einem Punkt enden, können nur in der Befehlssyntax erstellt werden. Variablen, die mit einem Punkt enden, können nicht in Dialogfeldern erstellt werden, die zum Erstellen neuer Variablen dienen.
- Variablennamen, die mit einem Unterstrich enden, sollten vermieden werden, da solche Namen mit den Namen von Variablen in Konflikt stehen können, die automatisch von Befehlen und Prozeduren erstellt werden.
- Reservierte Schlüsselwörter können nicht als Variablennamen verwendet werden. Reservierte Schlüsselwörter sind: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO und WITH.
- Variablennamen können aus einer beliebigen Kombination aus Klein- und Großbuchstaben bestehen. Die Groß- und Kleinschreibung bleibt auch bei der Anzeige erhalten.
- Wenn lange Variablennamen in der Ausgabe mehrere Zeilen einnehmen, erfolgt der Zeilenumbruch bei Unterstrichen, Punkten und dem Wechsel von Klein- zu Großschreibung.

Messniveau einer Variablen

Das Messniveau kann als metrische Skala (für numerische Daten in Form einer Intervall- oder Verhältnisskala), ordinal oder nominal angegeben werden. Nominale und ordinale Daten können entweder aus einem String (alphanumerisch) oder Zahlen bestehen.

- **Nominal.** Eine Variable kann als nominal behandelt werden, wenn ihre Kategorien sich nicht in eine natürliche Reihenfolge bringen lassen, z. B. die Firmenabteilung, in der eine Person arbeitet. Beispiele für nominale Variablen sind Region, Postleitzahl oder Religionszugehörigkeit.
- **Ordinal.** Eine Variable kann als ordinal behandelt werden, wenn ihre Werte für Kategorien stehen, die eine natürliche Reihenfolge aufweisen (z. B. Grad der Zufriedenheit mit Kategorien von sehr unzufrieden bis sehr zufrieden). Ordinale Variablen treten beispielsweise bei Einstellungsmessungen (Zufriedenheit oder Vertrauen) und bei Präferenzbeurteilungen auf.
- **Metrisch.** Eine Variable kann als metrisch behandelt werden, wenn ihre Werte geordnete Kategorien mit einer sinnvollen Metrik darstellen, sodass man sinnvolle Aussagen über die Abstände zwischen den Werten machen kann. Metrische Variablen sind beispielsweise Alter (in Jahren) oder Einkommen (in Geldeinheiten).

Anmerkung: Bei ordinalen String-Variablen wird angenommen, dass die Reihenfolge der Kategorien der alphabetischen Reihenfolge der String-Werte entspricht. Bei einer String-Variablen mit den Werten *Schwach*, *Mittel* und *Stark* werden die Kategorien beispielsweise in der Reihenfolge *Mittel*, *Schwach*, *Stark* und somit falsch angeordnet. Im allgemeinen ist die Verwendung von numerischem Code für ordinale Daten günstiger.

Neuen numerischen Variablen, die während einer Sitzung erstellt werden wird das metrische Messniveau zugewiesen. Für Daten, die aus Dateien in einem externen Dateiformat stammen, und für SPSS-Datendateien, die mit Programmversionen vor 8.0 erstellt wurden, wird eine Standardzuweisung für das Messniveau anhand folgender Regeln vorgenommen:

- Numerische Variablen mit weniger als 24 eindeutigen Werten und String-Variablen werden als nominal festgelegt.
- Numerische Variablen mit 24 oder mehr eindeutigen Werten werden als metrisch festgelegt.

Im Dialogfeld “Optionen” können Sie den Wert ändern, anhand dessen bei numerischen Variablen zwischen metrisch und nominal unterschieden wird. Für weitere Informationen siehe Optionen: Daten in Kapitel 45 auf S. 494.

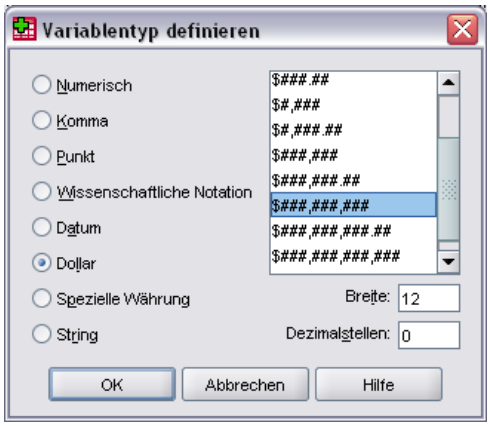
Mithilfe des Dialogfelds “Variableneigenschaften definieren”, das in dem Menü “Daten” verfügbar ist, können Sie das richtige Messniveau zuweisen. Für weitere Informationen siehe Zuweisen des Messniveaus in Kapitel 7 auf S. 114.

Variablentyp

Mit “Variablentyp definieren” wird für jede Variable der Datentyp angegeben. In der Standardeinstellung sind alle neuen Variablen als numerisch festgelegt. Mit “Variablentyp definieren” können Sie den Datentyp ändern. Dabei ist der Inhalt des Dialogfelds “Variablentyp definieren” von dem jeweils ausgewählten Datentyp abhängig. Bei einigen Datentypen gibt es

Textfelder für die Breite und die Anzahl der Dezimalstellen, bei anderen Datentypen können Sie einfach ein Format aus einer Liste mit Beispielen auswählen.

Abbildung 5-3
Dialogfeld "Variablentyp definieren"



Die folgenden Datentypen sind verfügbar:

Numerisch. Eine Variable, deren Werte Zahlen sind. Die Werte werden im numerischen Standardformat angezeigt. Numerische Wert können im Daten-Editor im Standardformat oder in wissenschaftlicher Notation eingegeben werden.

Komma. Eine numerische Variable, deren Werte mit Kommata als Tausender-Trennzeichen und Punkt als Dezimaltrennzeichen angezeigt werden. Numerische Werte für Kommavariablen können im Daten-Editor mit oder ohne Kommata oder in wissenschaftlicher Notation eingegeben werden. Die Werte können rechts neben dem Dezimaltrennzeichen kein Komma enthalten.

Punkt. Eine numerische Variable, deren Werte mit Punkten als Tausender-Trennzeichen und Komma als Dezimaltrennzeichen angezeigt werden. Numerische Werte für Punktvariablen können im Daten-Editor mit oder ohne Punkte oder in wissenschaftlicher Notation eingegeben werden. Die Werte können rechts neben dem Dezimaltrennzeichen keinen Punkt enthalten.

Wissenschaftliche Notation. Eine numerische Variable, deren Werte mit einem E und einer Zehnerpotenz mit Vorzeichen angezeigt werden. Numerische Werte für diese Variablen können im Daten-Editor mit oder ohne Potenz eingegeben werden. Dem Exponenten kann entweder ein E oder ein D (mit oder ohne Vorzeichen) oder ein Vorzeichen allein vorangestellt werden, beispielsweise 123, 1,23E2, 1,23D2, 1,23E+2 oder 1,23+2.

Datum. Eine numerische Variable, deren Werte in einem der Datums- oder Uhrzeitformate angezeigt werden. Wählen Sie ein Format aus der Liste aus. Sie können Datumsangaben mit Schrägstrichen, Bindestrichen, Punkten, Kommata oder Leerzeichen als Trennzeichen eingeben. Bei zweistelligen Jahresangaben hängt das Jahrhundert von den Einstellungen unter "Optionen" ab (wählen Sie dazu im Menü "Bearbeiten" den Befehl Optionen aus und klicken Sie dann auf die Registerkarte Daten).

Dollar. Eine numerische Variable mit führendem Dollarzeichen ($), deren Werte mit Kommata als Tausender-Trennzeichen und Punkt als Dezimaltrennzeichen angezeigt werden. Die Werte können mit und ohne das führende Dollarzeichen eingegeben werden.

Spezielle Währung. Eine numerische Variable, deren Werte in einem der benutzerdefinierten Währungsformate angezeigt wird, die im Dialogfeld "Optionen" auf der Registerkarte "Währung" definiert wurden. Zeichen, die in einem Währungsformat festgelegt wurden, können nicht für die Dateneingabe genutzt werden. Die Zeichen werden jedoch im Daten-Editor angezeigt.

String. Eine Variable, deren Werte nicht numerisch sind und die daher nicht in den Berechnungen verwendet werden. Die Werte dürfen beliebige Zeichen bis zur festgelegten Höchstlänge enthalten. Groß- und Kleinbuchstaben werden als separate Buchstaben betrachtet. Dieser Typ ist auch als alphanumerische Variable bekannt.

So definieren Sie einen Variablentyp:

- ▶ Klicken Sie auf die Schaltfläche in der Zelle *Typ* der Variablen, die Sie definieren möchten.
- ▶ Wählen Sie im Dialogfeld "Variablentyp definieren" den Datentyp aus.
- ▶ Klicken Sie auf OK.

Der Unterschied zwischen Eingabe- und Anzeigeformaten

Je nach Format kann die Anzeige von Werten in der Datenansicht von den eingegebenen und tatsächlich intern gespeicherten Werten abweichen. Im folgenden finden Sie einige allgemeine Richtlinien:

- In den Formaten numerisch, Komma und Punkt können Sie Werte mit jeder beliebigen Anzahl (bis zu 16) Dezimalstellen eingeben. Der gesamte Wert wird intern gespeichert. In der Datenansicht wird nur die definierte Anzahl der Stellen angezeigt. Werte mit mehr Dezimalstellen werden gerundet. In Berechnungen wird allerdings immer der vollständige Wert verwendet.
- Bei String-Variablen werden alle Werte bis zur maximalen Länge rechts mit Leerzeichen aufgefüllt. Bei einer String-Variablen mit einer maximalen Breite von 3 wird der Wert *Ja* intern als `'Ja '` gespeichert und ist somit nicht das gleiche wie `' Ja'`.
- Bei Datumsformaten können Sie Schrägstriche, Bindestriche, Leerzeichen, Kommata oder Punkte als Trennzeichen zwischen den Werten für Tag, Monat und Jahr verwenden. Für die Monatswerte können Sie Ziffern, Abkürzungen von drei Buchstaben Länge oder vollständige Namen eingeben. Datumsangaben im allgemeinen Format *tt-mmm-jj* werden mit Bindestrichen als Trennzeichen und mit aus drei Buchstaben bestehenden Abkürzungen für den Monat eingegeben. Datumsangaben im allgemeinen Format *tt-mm-jj* und *mm/tt/jj* werden mit Schrägstrichen als Trennzeichen und Zahlen für den Monat eingegeben. Die Daten werden intern als Anzahl der Sekunden gespeichert, die seit dem 14. Oktober 1582 vergangen sind. Den Jahrhundertbereich für zweistellige Jahresangaben können Sie in den Einstellungen unter "Optionen" angeben. Wählen Sie dazu im Menü "Bearbeiten" den Befehl Optionen und anschließend die Registerkarte Daten aus.
- In Zeitformaten können Sie Doppelpunkte, Punkte oder Leerzeichen als Trennzeichen zwischen Stunden, Minuten und Sekunden verwenden. Zeiten werden mit Doppelpunkten als Trennzeichen angezeigt. Intern werden Datumsangaben als Anzahl von Sekunden gespeichert, die ein Zeitintervall darstellen. So wird 10:00:00 beispielsweise intern als 36000, gespeichert, d. h. 60 (Sekunden pro Minute) x 60 (Minuten pro Stunde) x 10 (Stunden).

Variablenlabels

Sie können aussagekräftige Variablenlabels bis zu 256 Zeichen Länge (128 Zeichen für Double-Byte-Sprachen) zuweisen. Variablenlabels können Leerzeichen und reservierte Zeichen enthalten, die in Variablennamen nicht zulässig sind.

So legen Sie Variablenlabels fest:

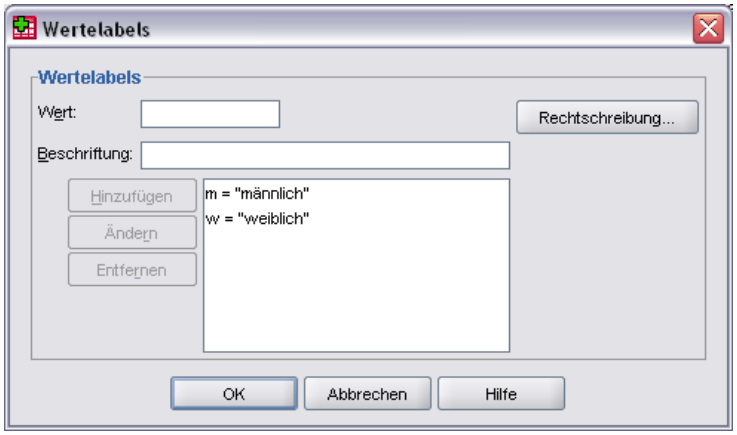
- ▶ Aktivieren Sie das Fenster des Daten-Editors.
- ▶ Doppelklicken Sie in der Datenansicht auf den Namen einer Variablen oben in einer Spalte oder klicken Sie auf die Registerkarte Variablenansicht.
- ▶ Geben Sie in der Zelle *Variablenlabel* für die Variable ein aussagekräftiges Variablenlabel ein.

Wertelabels

Sie können jedem Wert einer Variable ein beschreibendes Wertelabel zuordnen. Dies ist besonders nützlich, wenn Ihre Datendatei numerische Codes zur Darstellung nichtnumerischer Kategorien verwendet (zum Beispiel die Codes 1 und 2 für *Männlich* und *Weiblich*).

- Wertelabels können bis zu 120 Byte umfassen.

Abbildung 5-4
Dialogfeld "Wertelabels definieren"



So legen Sie Wertelabels fest:

- ▶ Klicken Sie auf die Schaltfläche in der Zelle *Wertelabels* der Variablen, die Sie definieren möchten.
- ▶ Geben Sie für jeden Wert den Wert und ein Label ein.
- ▶ Klicken Sie auf Hinzufügen, um das Wertelabel einzugeben.
- ▶ Klicken Sie auf OK.

Einfügen von Zeilenumbrüchen in Labels

Bei Variablenlabels und Wertelables werden in Pivot-Tabellen und Diagrammen automatisch Zeilenumbrüche eingefügt, wenn die Zelle bzw. der Bereich nicht breit genug ist, um das gesamte Label in einer Zeile anzuzeigen. Außerdem können Sie die Ergebnisse bearbeiten, um manuelle Zeilenumbrüche einzufügen, wenn der Label-Text an einer anderen Stelle umbrechen soll. Des Weiteren können Sie Variablenlabels und Wertelables erstellen, bei denen der Text *immer* an festgelegten Punkten umbricht und auf mehrere Zeilen verteilt angezeigt wird.

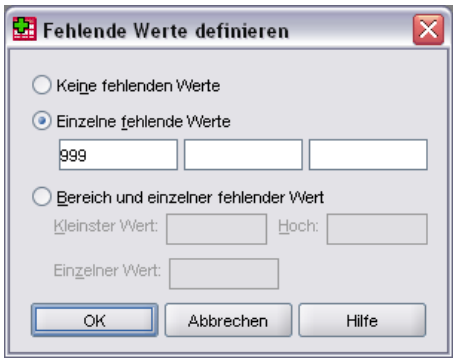
- Bei Variablenlabels wählen Sie in der Variablenansicht des Daten-Editors die Zelle *Variablenlabel* für die Variable aus.
- Bei Wertelabels wählen Sie in der Variablenansicht des Daten-Editors die Zelle *Wertelabels* für die Variable aus, klicken Sie auf die Schaltfläche in der Zelle und wählen Sie im Dialogfeld "Wertelabels definieren" das zu ändernde Label aus.
- Geben Sie an der Stelle, an der der Zeilenumbruch erfolgen soll, \n ein.

  Die Zeichenfolge \n wird in Pivot-Tabellen bzw. Diagrammen nicht angezeigt; sie wird als Zeichen für den Zeilenumbruch interpretiert.

Fehlende Werte

Mit der Option "Fehlende Werte" werden bestimmte Datenwerte als **benutzerdefiniert fehlende Werte** deklariert. So ist es zum Beispiel sinnvoll zu unterscheiden, ob Daten fehlen, weil eine befragte Person die Auskunft verweigerte oder weil die Frage sich nicht auf die befragte Person bezog. Datenwerte, die als benutzerdefiniert fehlende Werte angegeben sind, werden zur Sonderbehandlung gekennzeichnet und von den meisten Berechnungen ausgeschlossen.

Abbildung 5-5
Dialogfeld "Fehlende Werte definieren"



- Sie können entweder bis zu drei diskrete (einzelne) fehlende Werte, einen Bereich fehlender Werte oder einen Bereich und einen diskreten Wert eingeben.
- Bereiche können nur bei numerischen Variablen angegeben werden.
- Zunächst werden alle String-Variablen, einschließlich der Werte "Null" und "Leer", als gültig betrachtet, sofern diese nicht explizit als "fehlend" definiert worden sind.

- Fehlende Werte für String-Variablen dürfen nicht länger sein als 8 Byte. (Es gibt keine Obergrenze für die definierte Länge der String-Variablen, definierte fehlende Werte jedoch dürfen 8 Byte nicht überschreiten.)
- Wenn Sie String-Variablen mit den Werten "Null" oder "Leer" als fehlend definieren möchten, geben Sie in eines der Felder von Einzelne fehlende Werte ein einfaches Leerzeichen ein.

So definieren Sie fehlende Werte:

▶ Klicken Sie auf die Schaltfläche in der Zelle *Fehlende Werte* der Variablen, die Sie definieren möchten.

▶ Geben Sie die Werte oder den Bereich der Werte ein, welche die fehlenden Daten repräsentieren.

Spaltenbreite

Sie können die Spaltenbreite als Anzahl der angezeigten Zeichen festlegen. Die Spaltenbreite kann auch in der Datenansicht geändert werden, indem Sie auf eine Spaltenbegrenzung klicken und sie an die gewünschte Stelle ziehen.

Spaltenformate wirken sich nur auf die Anzeige der Werte im Daten-Editor aus. Änderungen der Spaltenbreite ändern nicht die definierte Länge einer Variablen.

Variablenausrichtung

Mit der Ausrichtung wird die Anzeige von Datenwerten und/oder Wertelabels in der Datenansicht festgelegt. Numerische Variablen werden in der Standardeinstellung rechts, String-Variablen links ausgerichtet. Diese Einstellung gilt nur für die Anzeige in der Datenansicht.

Zuweisen von Variablenattributen zu mehreren Variablen

Nachdem Sie Attribute zur Variablendefinition festgelegt haben, können Sie ein oder mehrere Attribute kopieren und einer oder mehreren Variablen zuweisen.

Das Zuweisen der Variablenattribute erfolgt über einfaches Kopieren und Einfügen. Sie verfügen über folgende Möglichkeiten:

- Einzelne Attribute (beispielsweise Wertelabels) können kopiert und in die gleiche Attributzelle einer oder mehrerer Variablen eingefügt werden.
- Alle Attribute einer Variablen können kopiert und in eine oder mehrere andere Variablen eingefügt werden.
- Mehrere neue Variablen mit sämtlichen Attributen einer kopierten Variable können erstellt werden.

Zuweisen von Variablenattributen zu anderen Variablen

So übertragen Sie einzelne Attribute aus einer bereits definierten Variablen:

- Wählen Sie in der Variablenansicht die Attributzelle aus, die Sie auf andere Variablen übertragen möchten.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Kopieren
- Wählen Sie die Attributzelle(n) aus, in die Sie das Attribut übertragen möchten. (Es können mehrere Zielvariablen ausgewählt werden.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Einfügen

  Wenn Sie die Attribute in leere Zeilen einfügen, werden neue Variablen erstellt, wobei allen Attributen mit Ausnahme des ausgewählten Attributs Standardwerte zugewiesen werden.

So übertragen Sie alle Attribute aus einer bereits definierten Variablen:

- Wählen Sie in der Variablenansicht die Zeilennummer der Variablen aus, deren Attribute Sie übertragen möchten. (Die gesamte Zeile wird markiert.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Kopieren
- Klicken Sie auf die Zeilennummer(n) der Variablen, der/denen Sie die Attribute zuweisen möchten. (Es können mehrere Zielvariablen ausgewählt werden.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Einfügen

Erstellen von mehreren neuen Variablen mit übereinstimmenden Attributen

- Klicken Sie in der Variablenansicht auf die Zeilennummer der Variablen, deren Attribute Sie auf die neue Variable übertragen möchten. (Die gesamte Zeile wird markiert.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Kopieren
- Klicken Sie auf die Nummer der leeren Zeile unterhalb der letzten definierten Variablen in der Datendatei.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Variablen einfügen...

- Geben Sie im Dialogfeld “Variablen einfügen” die Anzahl der Variablen ein, die Sie erstellen möchten.
- Geben Sie ein Präfix und eine Anfangsnummer für die neuen Variablen ein.
- Klicken Sie auf OK.

  Die Namen der neuen Variablen bestehen aus dem angegebenen Präfix und einer laufenden Nummer (ab der angegebenen Anfangsnummer).

Benutzerdefinierte Variablenattribute

Neben den standardmäßigen Variablenattributen (z. B. Wertelabels, fehlende Werte, Messniveau) können Sie Ihre eigenen benutzerdefinierten Variablenattribute erstellen. Wie die standardmäßigen Variablenattribute werden auch die benutzerdefinierten Attribute zusammen mit Datendateien im SPSS-Format gespeichert. Daher könnten Sie ein Variablenattribut erstellen, das den Antworttyp für Fragen in einer Umfrage (z. B. Einzelauswahl, Mehrfachauswahl, freie Antwort) oder die Formeln für berechnete Variablen identifiziert.

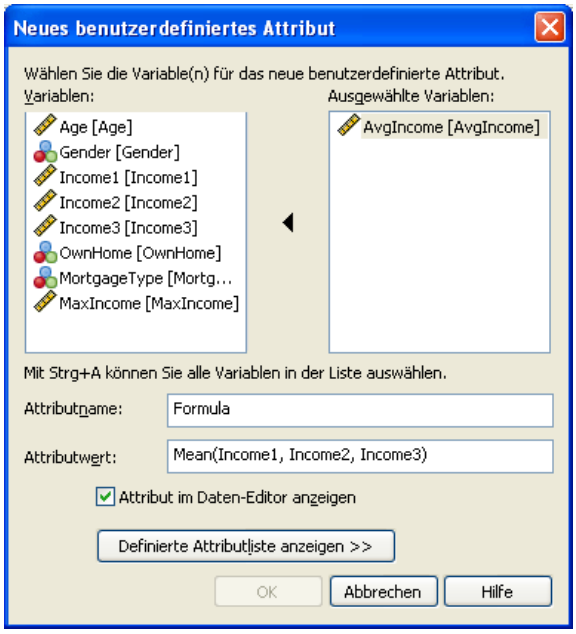
Erstellen von benutzerdefinierten Variablenattributen

So erstellen Sie neue benutzerdefinierte Attribute:

- Wählen Sie die folgenden Befehle aus den Menüs der Variablenansicht aus:

  Daten
  Neues benutzerdefiniertes Attribut...
- Ziehen Sie die Variablen, denen das neue Attribut zugewiesen werden soll, in die Liste “Ausgewählte Variablen” und legen Sie sie dort ab.
- Geben Sie einen Namen für das Attribut ein. Für Attributnamen gelten dieselben Regeln wie für Variablennamen. Für weitere Informationen siehe Variablennamen auf S. 83.
- Geben Sie einen optionalen Wert für das Attribut ein. Wenn Sie mehrere Variablen auswählen, wird der Wert allen ausgewählten Variablen zugewiesen. Sie können dieses Feld leer lassen und anschließend in der Variablenansicht Werte für die einzelnen Variablen eingeben.

Abbildung 5-6
Dialogfeld "Neues benutzerdefiniertes Attribut"



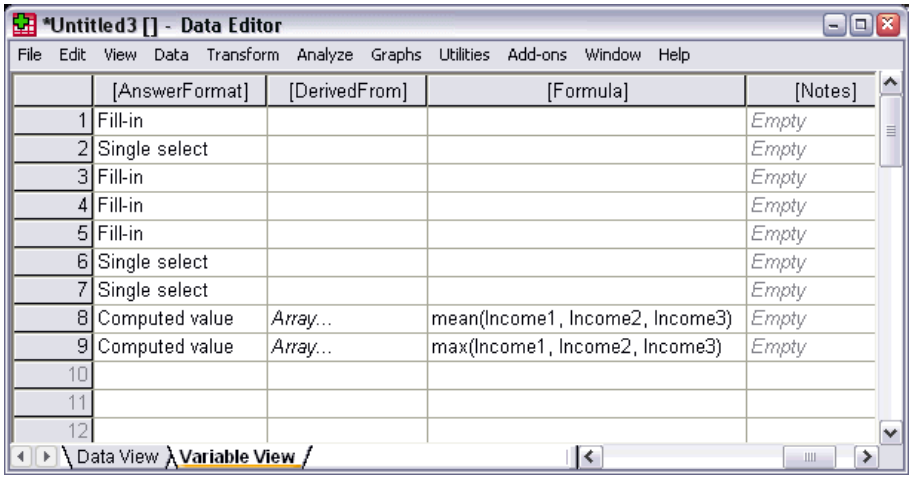
Attribut im Daten-Editor anzeigen. Zeigt das Attribut in der Variablenansicht des Daten-Editors an. Weitere Informationen zur Steuerung der Anzeige benutzerdefinierter Attribute finden Sie unter Anzeigen und Bearbeiten von benutzerdefinierten Variablenattributen.

Definierte Attributliste anzeigen. Zeigt eine Liste benutzerdefinierter Attribute an, die bereits für das Daten-Set definiert wurden. Attributnamen, die mit einem Dollarzeichen ($) beginnen, sind reservierte Attribute, die nicht geändert werden können.

Anzeigen und Bearbeiten von benutzerdefinierten Variablenattributen

Benutzerdefinierte Variablenattribute können im Daten-Editor in der Variablenanzeige angezeigt und bearbeitet werden.

Abbildung 5-7
In der Variablenansicht anzgezeigte benutzerdefinierte Attribute



- Die Namen benutzerdefinierter Variablenattribute sind in eckige Klammern eingeschlossen.
- Attributnamen, die mit einem Dollarzeichen beginnen, sind reserviert und können nicht geändert werden.
- Eine leere Zelle zeigt an, dass das Attribut für die betreffende Variable nicht vorhanden ist. Wenn der Text *Leer* in einer Zelle angezeigt wird, bedeutet dies, dass das Attribut für die betreffende Variable vorhanden ist, ihm jedoch kein Wert zugewiesen wurde. Sobald Sie Text in die Zelle eingeben, ist das Attribut mit dem von Ihnen eingegebenen Wert für die betreffende Variable vorhanden.
- Wenn der Text *Array...* in einer Zelle angezeigt wird, bedeutet dies, dass es sich dabei um ein Attribut-Array handelt, ein Attribut, das mehrere Werte enthält. Klicken Sie auf die Schaltfläche in der Zelle, um die Liste der Werte anzuzeigen.

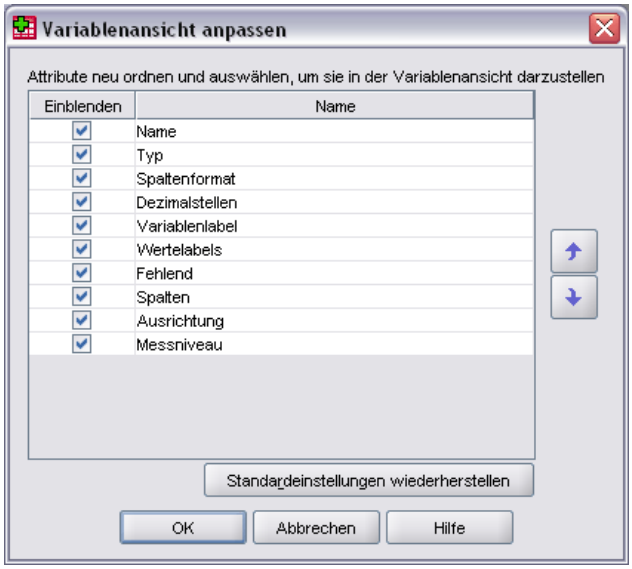
So können Sie benutzerdefinierte Variablenattribute anzeigen und bearbeiten:

▶ Wählen Sie die folgenden Befehle aus den Menüs der Variablenansicht aus:

Ansicht
  Variablenansicht anpassen...

▶ Aktivieren Sie die benutzerdefinierten Variablenattribute, die angezeigt werden sollen. (Die benutzerdefinierten Variablenattribute sind in eckige Klammern eingeschlossen.)

Abbildung 5-8
Variablenansicht anpassen

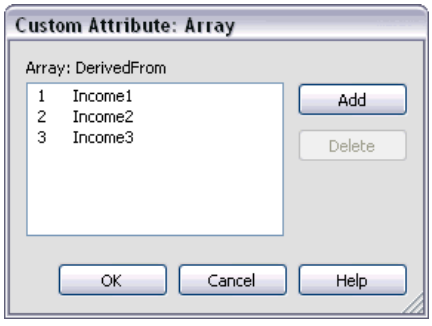


Sobald die Attribute in der Variablenansicht angezeigt werden, können Sie sie direkt im Daten-Editor bearbeiten.

Variablenattribut-Arrays

Der Text *Array...* – in einer Zelle für ein benutzerdefiniertes Variablenattribut in der Variablenansicht oder im Dialogfeld "Benutzerdefinierte Variableneigenschaften" unter "Variableneigenschaften definieren" angezeigt – bedeutet, dass es sich um ein **Attribut-Array** handelt, ein Attribut, das mehrere Werte enthält. Sie könnten beispielsweise über ein Attribut-Array verfügen, mit dem alle Quellvariablen identifiziert werden, die zur Berechnung einer abgeleiteten Variablen verwendet werden. Klicken Sie auf die Schaltfläche in der Zelle, um die Liste der Werte anzuzeigen und zu bearbeiten.

Abbildung 5-9
Dialogfeld "Benutzerdefiniertes Attribut-Array"



Anpassen der Variablenansicht

Mit “Variablenansicht anpassen” können Sie festlegen, welche Attribute in der Variablenansicht angezeigt werden (z. B. Name, Typ, Variablenlabel) und in welcher Reihenfolge diese Attribute angezeigt werden.

- Alle benutzerdefinierten Variablenattribute, die mit dem Daten-Set verknüpft sind, sind in eckige Klammern eingeschlossen. Für weitere Informationen siehe Erstellen von benutzerdefinierten Variablenattributen auf S. 91.
- Benutzerdefinierte Anzeigeeinstellungen werden bei Datendateien im SPSS-Format mitgespeichert.
- Außerdem können Sie die Standardvorgaben für Anzeige und Reihenfolge der Attribute in der Variablenansicht festlegen. Für weitere Informationen siehe Ändern der Standard-Variablenansicht in Kapitel 45 auf S. 496.

So passen Sie die Variablenansicht an:

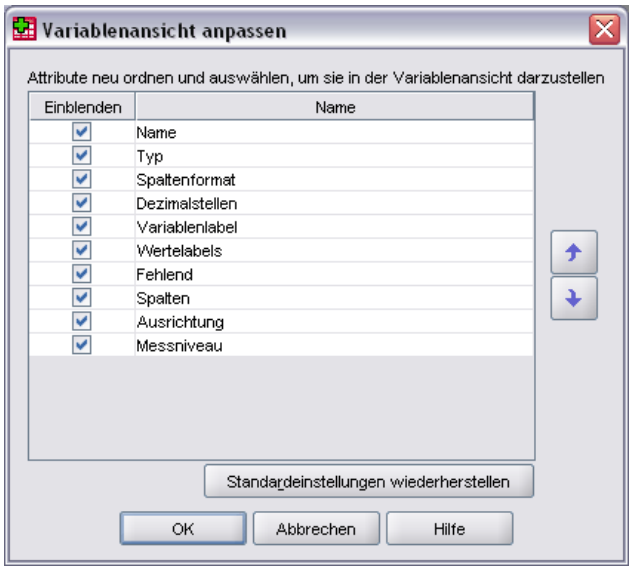
▶ Wählen Sie die folgenden Befehle aus den Menüs der Variablenansicht aus:

Ansicht
  Variablenansicht anpassen...

▶ Aktivieren Sie die Variablenattribute, die angezeigt werden sollen.

▶ Die Anzeigereihenfolge der Attribute können Sie mit der nach oben bzw. nach unten weisenden Pfeilschaltfläche ändern.

Abbildung 5-10
Dialogfeld “Variablenansicht anpassen”



Standardeinstellungen wiederherstellen. Übernimmt die Standardeinstellungen für Anzeige und Reihenfolge.

Rechtschreibprüfung bei Variablen- und Wertelabels

So können Sie die Rechtschreibung von Variablen- und Wertelabels überprüfen:

▶ Klicken Sie in der Variablenansicht auf die Spalte *Variablenlabel* bzw. *Wertelabels* und wählen folgende Option im Kontextmenü:
Rechtschreibung

oder

▶ Wählen Sie die folgenden Befehle aus den Menüs der Variablenansicht aus:
Extras
Rechtschreibung

oder

▶ Klicken Sie im Dialogfeld "Wertelabels definieren" auf Rechtschreibung. (Dadurch wird die Rechtschreibprüfung auf die Wertelabels für eine bestimmte Variable beschränkt.)

Die Rechtschreibprüfung ist auf Variablen- und Wertelabels in der Variablenansicht des Daten-Editors beschränkt.

Eingeben von Daten

In der Datenansicht können Sie die Daten direkt in den Daten-Editor eingeben. Sie können Daten in beliebiger Reihenfolge eingeben. Sie können Daten nach Fall oder Variable, für ausgewählte Bereiche oder einzelne Zellen eingeben.

- Die aktive Zelle wird hervorgehoben.
- In der linken oberen Ecke des Daten-Editors werden der Name der Variablen und die Zeilennummer der aktiven Zelle angezeigt.
- Wenn Sie eine Zelle auswählen und einen Datenwert eingeben, wird der Wert im Zellen-Editor am oberen Rand des Daten-Editors angezeigt.
- Datenwerte werden nicht aufgezeichnet, bis Sie die Eingabetaste drücken oder eine andere Zelle wählen.
- Wenn Sie andere Daten als einfache numerische Daten eingeben möchten, müssen Sie zuerst den Variablentyp definieren.

Wenn Sie einen Wert in eine leere Spalte eingeben, wird vom Daten-Editor automatisch eine neue Variable erstellt und ein neuer Variablenname zugewiesen.

So geben Sie numerische Daten ein:

▶ Wählen Sie in der Datenansicht eine Zelle aus.

▶ Geben Sie den Datenwert ein. (Der Wert wird im Zellen-Editor am oberen Rand des Daten-Editors angezeigt.)

▶ Drücken Sie die Eingabetaste oder wählen Sie eine andere Zelle aus, um den Wert aufzuzeichnen.

So geben Sie nichtnumerische Daten ein:

- ▶ Doppelklicken Sie in der Datenansicht auf den Namen einer Variablen oben in einer Spalte oder klicken Sie auf die Registerkarte Variablenansicht.
- ▶ Klicken Sie auf die Schaltfläche in der Zelle *Typ* der Variablen.
- ▶ Wählen Sie im Dialogfeld “Variablentyp definieren” den Datentyp aus.
- ▶ Klicken Sie auf OK.
- ▶ Doppelklicken Sie auf die Zeilennummer oder klicken Sie auf die Registerkarte Datenansicht.
- ▶ Geben Sie die Daten in die Spalte für die neu definierte Variable ein.

So verwenden Sie Wertelabels bei der Dateneingabe:

- ▶ Falls in der Datenansicht gegenwärtig keine Wertelabels angezeigt werden, wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Wertelabels

- ▶ Klicken Sie auf die Zelle, in der Sie den Wert eingeben möchten.
- ▶ Wählen Sie aus der Dropdown-Liste ein Wertelabel aus.

  Der Wert wird eingegeben und das Wertelabel in der Zelle angezeigt.

  Anmerkung: Dies ist jedoch nur möglich, wenn für die Variable Wertelabels definiert wurden.

Einschränkungen für die Datenwerte im Daten-Editor

Der definierte Variablentyp und die definierte Variablenlänge bestimmen, welche Werte in der Datenansicht in die Zelle eingegeben werden können.

- Wenn Sie ein für den definierten Variablentyp nicht zugelassenes Zeichen eingeben, wird die Eingabe nicht angenommen.
- Bei String-Variablen sind nicht mehr Zeichen erlaubt, als die definierte Länge zulässt.
- Bei numerischen Variablen können ganzzahlige Werte eingegeben werden, welche die definierte Länge überschreiten, aber der Daten-Editor zeigt entweder die Werte in wissenschaftlicher Notation oder einen Teil des Werts gefolgt von Auslassungszeichen (...) an, um zu kennzeichnen, dass der Wert länger als zulässig ist. Um den Wert in der Zelle anzuzeigen, müssen Sie die definierte Länge der Variablen ändern.

  Anmerkung: Das Ändern der Spaltenbreite hat keinen Einfluss auf die Länge der Variablen.

Bearbeiten von Daten

Mit dem Daten-Editor können Sie die Datenwerte in der Datenansicht auf verschiedene Arten bearbeiten. Sie verfügen über folgende Möglichkeiten:

- Datenwerte ändern

- Datenwerte ausschneiden, kopieren und einfügen
- Fälle hinzufügen und löschen
- Variablen hinzufügen und löschen
- Reihenfolge der Variablen ändern

Ersetzen oder Ändern von Datenwerten

So löschen Sie den alten Wert und geben einen neuen Wert ein:

- ▶ Doppelklicken Sie in der Datenansicht auf die Zelle. (Der Wert der Zelle wird im Zellen-Editor angezeigt.)
- ▶ Bearbeiten Sie den Wert direkt in der Zelle oder im Zellen-Editor.
- ▶ Drücken Sie die Eingabetaste oder wählen Sie eine andere Zelle zum Aufzeichnen des neuen Werts.

Ausschneiden, Kopieren und Einfügen von Datenwerten

Im Daten-Editor können Sie einzelne Werte aus Zellen oder Gruppen von Werten ausschneiden, kopieren und einfügen. Sie verfügen über folgende Möglichkeiten:

- Einen einzelnen Zellenwert in eine andere Zelle verschieben oder kopieren
- Einen einzelnen Zellenwert in eine Gruppe von Zellen verschieben oder kopieren
- Die Werte für einen einzelnen Fall (Zeile) in mehrere Fälle verschieben oder kopieren
- Die Werte für eine einzelne Variable (Spalte) in mehrere Variablen verschieben oder kopieren
- Eine Gruppe von Zellenwerten in eine andere Gruppe von Zellen verschieben oder kopieren

Umwandlung für eingefügte Werte im Daten-Editor

Wenn die definierten Variablentypen der Quell- und Zielzellen nicht übereinstimmen, versucht der Daten-Editor, den Wert zu konvertieren. Wenn eine Umwandlung nicht möglich ist, wird der systemdefiniert fehlende Wert in die Zielzelle eingefügt.

Umwandlung von numerischen Formaten oder Datumsformaten in Strings. Numerische Formate (zum Beispiel numerisch, Dollar, Punkt oder Komma) und Datumsformate werden zu Strings konvertiert, wenn sie in eine Zelle für String-Variablen eingefügt werden. Der Stringwert ist der in der Zelle angezeigte numerische Wert. So wird zum Beispiel bei einer Variablen im Dollarformat das angezeigte Dollarzeichen zum Bestandteil des Stringwerts. Werte, welche die definierte Länge der String-Variablen übersteigen, werden abgeschnitten.

Umwandlung von Strings in numerische Werte oder Datumswerte. Stringwerte, die akzeptable Zeichen für das numerische Format oder das Datumsformat der Zielzelle enthalten, werden in den äquivalenten numerischen Wert oder Datumswert konvertiert. So wird zum Beispiel der String-Wert “25/12/91” in ein gültiges Datum konvertiert, falls das Format der Zielzelle vom Typ Tag-Monat-Jahr ist. Falls das Format der Zielzelle aber vom Typ Monat-Tag-Jahr ist, dann wird er in den systemdefiniert fehlenden Wert konvertiert.

Umwandlung von Datumswerten in numerische Werte. Werte für Datum und Uhrzeit werden in eine Anzahl von Sekunden umgewandelt, wenn die Zielzelle im numerischen Format ist (zum Beispiel numerisch, Dollar, Punkt oder Komma). Da Datumsangaben intern als die Anzahl der seit dem 14. Oktober 1582 vergangenen Sekunden gespeichert werden, kann das Umwandeln von Daten in numerische Werte zu extrem großen Zahlen führen. Das Datum 10/29/91 wird beispielsweise in den numerischen Wert 12.908.073.600 umgewandelt.

Umwandlung von numerischen Werten in Datums- oder Uhrzeitangaben. Numerische Werte werden in Datums- oder Uhrzeitangaben umgewandelt, wenn der Wert eine Anzahl von Sekunden darstellt, die in eine gültige Uhrzeit- oder Datumsangabe umgewandelt werden kann. Bei Datumsangaben werden numerische Werte unter 86.400 in den systemdefiniert fehlenden Wert umgewandelt.

Einfügen von neuen Fällen

Durch die Eingabe von Daten in eine Zelle in einer leeren Zeile wird automatisch ein neuer Fall angelegt. Für alle anderen Variablen dieses Falls fügt der Daten-Editor den systemdefiniert fehlenden Wert ein. Wenn sich zwischen dem neuen Fall und den bereits vorhandenen Fällen Leerzeilen befinden, dann werden die Leerzeilen ebenfalls neue Fälle mit dem systemdefiniert fehlenden Wert für alle Variablen. Sie können neue Fälle auch zwischen vorhandenen Fällen einfügen.

So fügen Sie einen neuen Fall zwischen vorhandenen Fällen ein:

- Wählen Sie in der Datenansicht eine Zelle in dem Fall (in der Zeile) unterhalb der Position aus, an der Sie den neuen Fall einfügen möchten.

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Fälle einfügen

  Für den Fall wird eine neue Zeile eingefügt und alle Variablen erhalten den systemdefiniert fehlenden Wert.

Einfügen von neuen Variablen

Wenn Sie in der Datenansicht in eine leere Spalte oder in der Variablenansicht in eine leere Zeile Daten eingeben, wird automatisch eine neue Variable erstellt. Dieser Variable wird ein Standardname (Präfix *var* und eine laufende Nummer) und ein Standard-Datentyp (numerisch) zugewiesen. Der Daten-Editor fügt für alle Fälle der neuen Variable den systemdefiniert fehlenden Wert ein. Wenn in der Datenansicht leere Spalten oder in der Variablenansicht leere Zeilen zwischen der neuen Variablen und bereits vorhandenen Variablen stehen, werden diese Zeilen und Spalten ebenfalls in neue Variablen umgewandelt. Auch diesen Variablen wird der systemdefiniert fehlende Wert für alle Fälle zugewiesen. Sie können neue Variablen auch zwischen vorhandenen Variablen einfügen.

So fügen Sie eine neue Variable zwischen vorhandenen Variablen ein:

- Wählen Sie eine beliebige Zelle rechts neben (in der Datenansicht) oder direkt unter (in der Variablenansicht) der Position aus, an der Sie die neue Variable einfügen möchten.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Variable einfügen

  Neue Variablen werden mit dem systemdefiniert fehlenden Wert für alle Fälle eingefügt.

So verschieben Sie Variablen:

- Um die Variable auszuwählen, klicken Sie in der Datenansicht auf den Variablennamen oder in der Variablenansicht auf die Zeilennummer für die Variable.
- Ziehen Sie die Variable an die neue Position.
- Wenn Sie die Variable zwischen zwei vorhandenen Variablen einfügen möchten: Legen Sie in der Datenansicht die Variable in der Variablenspalte rechts neben der Stelle ab, an der Sie die Variable platzieren möchten, oder legen Sie die Variable in der Variablenansicht in der Variablenzeile unterhalb der Stelle ab, an der Sie die Variable platzieren möchten.

So ändern Sie den Datentyp:

Sie können den Datentyp einer Variablen jederzeit ändern, indem Sie das Dialogfeld "Variablentyp definieren" in der Variablenansicht verwenden. Der Daten-Editor versucht, bereits vorhandene Werte in den neuen Typ zu konvertieren. Ist eine Umwandlung nicht möglich, wird der systemdefiniert fehlende Wert zugewiesen. Die Umwandlungsregeln sind dieselben wie beim Einfügen von Datenwerten in eine Variable mit anderem Formattyp. Falls die Änderung des Datenformats zum Verlust von Definitionen fehlender Werte oder von Wertelabels führen könnte, zeigt der Daten-Editor eine Warnung an und Sie werden gefragt, ob die Änderung trotzdem durchgeführt werden soll.

Suchen von Fällen bzw. Variablen

Im Dialogfeld "Gehe zu Fall/Variable" können Sie den Fall (die Zeile) mit der angegebenen Nummer bzw. dem angegebenen Variablennamen suchen.

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Gehe zu Fall

  oder

  Bearbeiten
  Gehe zu Variable

- Geben Sie bei Fällen in einen ganzzahligen Wert ein, der die aktuelle Zeilennummer in der Datenansicht darstellt.

Anmerkung: Die aktuelle Zeilennummer eines Falls kann sich durch Sortierungen und andere Aktionen ändern.

oder

- Geben Sie bei Variablen den Variablennamen ein oder wählen Sie die Variable aus der Dropdown-Liste aus.

Abbildung 5-11
Dialogfeld "Gehe zu"

Suchen und Ersetzen von Daten- und Attributwerten

So können Sie Datenwerte in der Datenansicht bzw. Attributwerte in der Variablenansicht suchen bzw. ersetzen:

- Klicken Sie auf eine Zelle in der Spalte, die durchsucht werden soll. (Das Suchen und Ersetzen von Werten ist auf eine einzelne Spalte beschränkt.)
- Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Suchen

oder

Bearbeiten
  Ersetzen

Datenansicht

- Eine Aufwärtssuche ist in der Datenansicht nicht möglich. Die Suchrichtung ist stets abwärts.
- Bei Datums- und Zeitangaben werden die formatierten Werte so gesucht, wie der Datenansicht angezeigt. So wird beispielsweise ein als "10/28/2007" angezeigtes Datum bei einer Suche nach dem Datum "10-28-2007" nicht gefunden.
- Bei anderen numerischen Variablen werden mit Enthält, Beginnt mit und Endet mit formatierte Werte gesucht. So wird beispielsweise bei der Option Beginnt mit mit dem Suchwert "$123" für eine Variable im Dollarformat sowohl $123,00 als auch $123,40 gefunden, nicht jedoch $1.234. Bei der Option Gesamte Zelle kann der Suchwert formatiert oder unformatiert sein (einfaches numerisches Format), es werden jedoch nur exakte numerische Werte (mit der im Daten-Editor angezeigten Genauigkeit) als Abhängige Variable ausgegeben.

- Wenn für die ausgewählte Variablenspalte keine Wertelabels angezeigt werden, wird der Labeltext durchsucht (nicht der zugrunde liegenden Datenwert) und der Text des Labels kann nicht ersetzt werden.

Variablenansicht

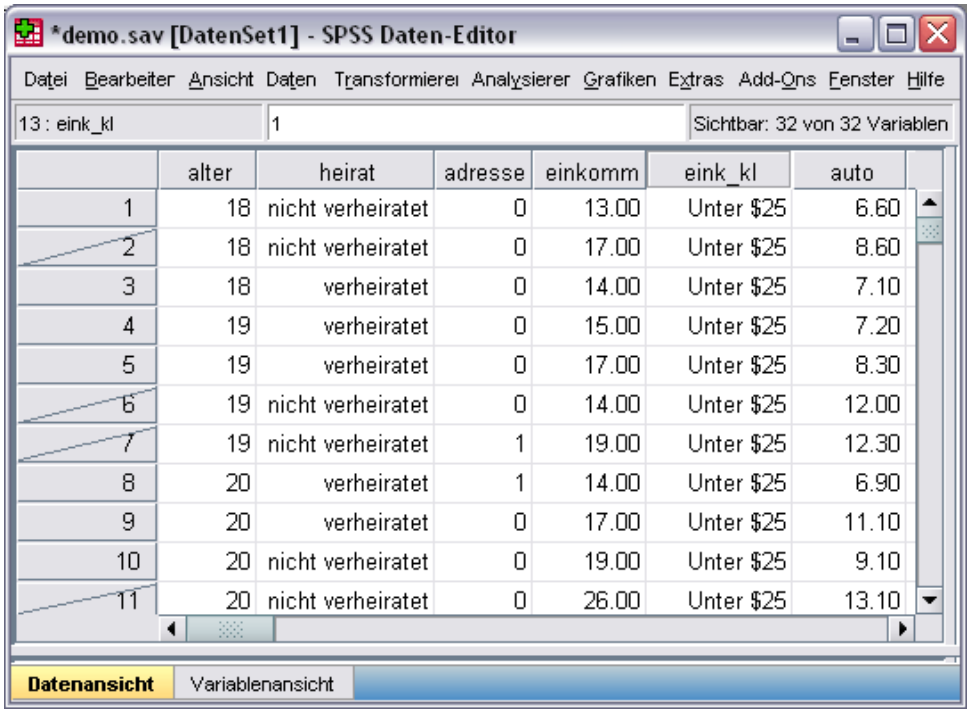
- "Suchen" steht nur für die Spalten *Name*, *Variablenlabel*, *Wertelabels*, *Fehlende Werte* und für Spalten mit benutzerdefinierten Variablenattributen zur Verfügung.
- "Ersetzen" steht nur für die Spalten *Variablenlabel*, *Wertelabels* und für Spalten mit benutzerdefinierten Variablenattributen zur Verfügung.
- In der Spalte *Wertelabels* kann der Suchstring entweder mit dem Datenwert oder mit einem Wertelabel übereinstimmen.

  Anmerkung: Durch das Ersetzen des Datenwerts werden alle vorherigen Wertelabel gelöscht, die dem betreffenden Wert zugeordnet waren.

Status für die Fallauswahl im Daten-Editor

Wenn Sie eine Teilmenge von Fällen ausgewählt, nicht ausgewählte Fälle jedoch nicht verworfen haben, sind die nicht ausgewählten Fälle im Daten-Editor mit einer diagonalen Linie (Schrägstrich) durch die Zeilennummer gekennzeichnet.

Abbildung 5-12
Gefilterte Fälle im Daten-Editor



	alter	heirat	adresse	einkomm	eink_kl	auto
1	18	nicht verheiratet	0	13.00	Unter $25	6.60
2	18	nicht verheiratet	0	17.00	Unter $25	8.60
3	18	verheiratet	0	14.00	Unter $25	7.10
4	19	verheiratet	0	15.00	Unter $25	7.20
5	19	verheiratet	0	17.00	Unter $25	8.30
6	19	nicht verheiratet	0	14.00	Unter $25	12.00
7	19	nicht verheiratet	1	19.00	Unter $25	12.30
8	20	verheiratet	1	14.00	Unter $25	6.90
9	20	verheiratet	0	17.00	Unter $25	11.10
10	20	nicht verheiratet	0	19.00	Unter $25	9.10
11	20	nicht verheiratet	0	26.00	Unter $25	13.10

Optionen für die Anzeige im Daten-Editor

Das Menü “Ansicht” bietet verschiedene Anzeigeoptionen für den Daten-Editor:

Schriftarten. Mit dieser Option können Sie die Schrifteigenschaften der Datenanzeige festlegen.

Gitterlinien. Mit dieser Option werden Gitterlinien ein- und ausgeblendet.

Wertelabels. Mit dieser Option wechseln Sie zwischen der Anzeige der tatsächlichen Datenwerte und der benutzerdefinierten beschreibenden Wertelabels hin und her. Diese Option ist nur in der Datenansicht verfügbar.

Verwenden mehrerer Ansichten

In der Datenansicht können Sie mehrere Ansichten (Fensterbereiche) mithilfe der Fensterteiler unterhalb der horizontalen Bildlaufleiste und rechts neben der vertikalen Bildlaufleiste anlegen.

Des weiteren können Sie Fensterteiler über das Menü “Fenster” einfügen und wieder entfernen. So fügen Sie Fensterteiler ein:

▶ Wählen Sie die folgenden Befehle aus den Menüs der Datenansicht aus:

Fenster
  Aufteilen

Fensterteiler werden oberhalb und links von der ausgewählten Zelle eingefügt.

- Wenn Sie die obere linke Zelle auswählen, werden die Fensterteiler so platziert, dass die aktuelle Ansicht horizontal und vertikal etwa gleich geteilt wird.
- Wählen Sie nicht die oberste Zelle in der ersten Spalte aus, wird ein horizontaler Fensterteiler oberhalb der ausgewählten Zelle eingefügt.
- Wählen Sie nicht die erste Zelle in der obersten Zeile aus, wird ein vertikaler Fensterteiler links neben der ausgewählten Zelle eingefügt.

Drucken aus dem Daten-Editor

Eine Datendatei wird so gedruckt, wie sie auf dem Bildschirm angezeigt wird.

- Die in der gegenwärtig angezeigten Ansicht enthaltenen Informationen werden gedruckt. In der Datenansicht werden die Daten gedruckt. In der Variablenansicht werden die Informationen zu den Variablendefinitionen gedruckt.
- Gitterlinien werden gedruckt, wenn sie gegenwärtig in der ausgewählten Ansicht angezeigt werden.
- Wenn in der Datenansicht gegenwärtig Wertelabels angezeigt werden, werden diese gedruckt. Andernfalls werden die eigentlichen Datenwerte gedruckt.

Verwenden Sie das Menü “Ansicht” im Fenster des Daten-Editors, um Gitterlinien ein- oder auszublenden und zwischen der Anzeige von Datenwerten bzw. Wertelabels umzuschalten.

So drucken Sie den Inhalt des Daten-Editors:

▶ Aktivieren Sie das Fenster des Daten-Editors.

- Klicken Sie auf die Registerkarte der Ansicht, die gedruckt werden soll.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Datei
  Drucken...

Kapitel 6

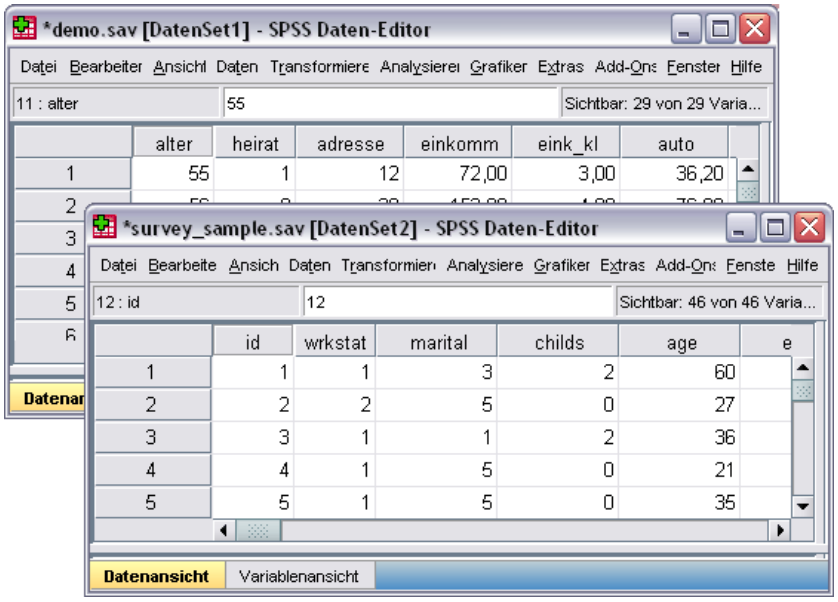
Arbeiten mit mehreren Datenquellen

Ab Version 14.0 können mehrere Datenquellen gleichzeitig geöffnet sein. Dies vereinfacht Folgendes:

- Wechseln zwischen Datenquellen
- Vergleichen der verschiedenen Datenquellen
- Kopieren und Einfügen von Daten zwischen Datenquellen
- Erstellen von Teilmengen der Fälle und/oder Variablen für die Analyse
- Zusammenführen von verschiedenen Datenquellen mit unterschiedlichen Datenformaten (beispielsweise Tabellenkalkulationsblätter, Datenbanken, Textdaten), ohne dass zuerst jede Datenquelle gespeichert werden muss.

Grundsätzlicher Umgang mit mehreren Datenquellen

Abbildung 6-1
Zwei gleichzeitig geöffnete Datenquellen

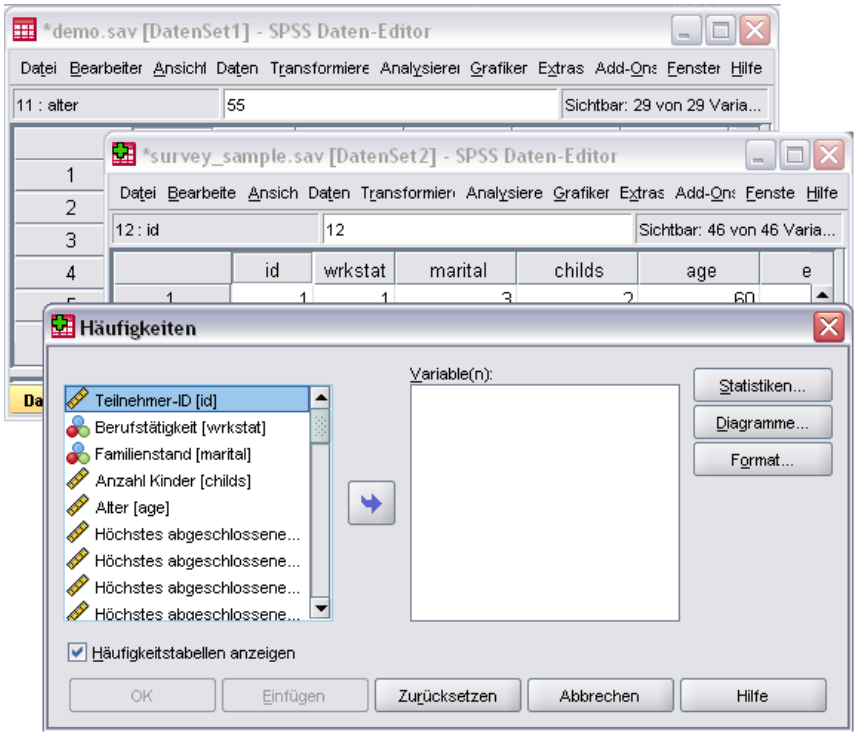


Standardmäßig wird jede Datenquelle, die Sie öffnen, in einem neuen Fenster des Daten-Editors angezeigt. (Unter Optionen: Allgemein finden Sie Informationen zum Ändern des Standardverhaltens, sodass nur jeweils ein einziges Daten-Set in einem einzigen Daten-Editor-Fenster angezeigt wird.)

- Jede zuvor geöffnete Datenquelle bleibt geöffnet und für die weitere Verwendung verfügbar.
- Wenn Sie eine Datenquelle öffnen, wird diese automatisch zur **Arbeitsdatei**.
- Sie können die Arbeitsdatei wechseln, indem Sie auf eine beliebige Stelle im Fenster "Daten-Editor" der gewünschten Datenquelle klicken oder das Fenster "Daten-Editor" für diese Datenquelle aus dem Menü "Fenster" auswählen.

- Nur die Variablen in der Arbeitsdatei sind für die Analyse verfügbar.

Abbildung 6-2
Variablenliste mit Variablen aus der Arbeitsdatei



- Sie können die Arbeitsdatei nicht wechseln, wenn ein Dialogfeld geöffnet ist, mit dem auf die Daten zugegriffen wird (einschließlich aller Dialogfelder, in denen Variablenlisten angezeigt werden).
- Während einer Sitzung muss mindestens ein Fenster des Daten-Editors geöffnet sein. Wenn Sie das letzte geöffnete Fenster des Daten-Editors schließen, wird SPSS automatisch beendet. Dabei werden Sie aufgefordert, die Änderungen zu speichern.

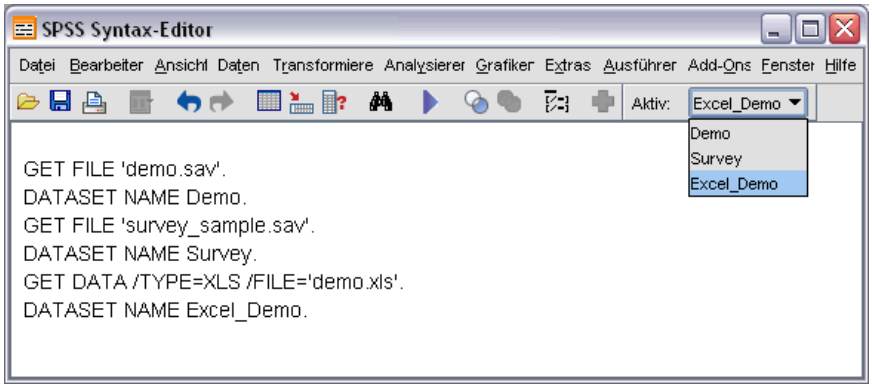
Arbeiten mit mehreren Daten-Sets in der Befehlssyntax

Wenn Sie die Befehlssyntax verwenden, um Datenquellen zu öffnen (z. B. `GET FILE` oder `GET DATA`), muss mithilfe von `DATASET NAME` jedes Daten-Set eindeutig benannt werden, damit mehr als eine Datenquelle gleichzeitig geöffnet werden kann.

Bei der Arbeit mit der Befehlssyntax wird der Name des aktiven Daten-Sets (Arbeitsdatei) in der Symbolleiste des Syntaxfensters angezeigt. Mit den folgenden Aktionen kann die Arbeitsdatei geändert werden:

- Über den Befehl `DATASET ACTIVATE`.
- Klicken Sie an eine beliebige Stelle im Fenster “Daten-Editor” eines Daten-Sets.
- Auswählen eines Daten-Set-Namens in der Symbolleiste im Syntaxfenster.

Abbildung 6-3
Offene Daten-Sets, die in der Symbolleiste des Syntaxfensters angezeigt werden.



Kopieren und Einfügen von Informationen zwischen Daten-Sets

Sowohl Daten als auch die Attribute zur Variablendefinition werden grundsätzlich auf die gleiche Art aus einem Daten-Set in ein anderes Daten-Set kopieren, in der Sie Informationen innerhalb einer einzelnen Datendatei kopieren und einfügen.

- Beim Kopieren und Einfügen von ausgewählten Datenzellen in der Datenansicht werden nur die Datenwerte eingefügt, nicht die Attribute zur Variablendefinition.
- Beim Kopieren und Einfügen einer vollständigen Variablen in der Datenansicht durch Auswählen der Variablen im Spaltenkopf werden alle Daten und alle Attribute zur Variablendefinition für diese Variable eingefügt.
- Beim Kopieren und Einfügen von Attributen zur Variablendefinition oder vollständigen Variablen in der Variablenansicht werden die ausgewählten Attribute (oder die vollständige Variablendefinition) eingefügt, nicht jedoch die Datenwerte.

Umbenennen von Daten-Sets

Wenn Sie eine Datenquelle über die Menüs und Dialogfelder öffnen, wird jeder Datenquelle automatisch der Name *DatenSetn* zugewiesen, wobei *n* eine fortlaufende ganze Zahl ist. Wenn Sie eine Datenquelle mit der Befehlssyntax öffnen, wird dem Daten-Set kein Name zugewiesen, sofern Sie nicht ausdrücklich mit `DATASET NAME` einen Namen angeben. So vergeben Sie aussagekräftigere Namen für die Daten-Sets:

▶ Wählen Sie die folgenden Optionen für das Daten-Set, dessen Namen Sie ändern möchten, aus den Menüs im Fenster des Daten-Editors aus:

Datei
  Daten-Set umbenennen...

▶ Geben Sie einen neuen Namen für das Daten-Set ein, der den Regeln für Variablennamen entspricht. Für weitere Informationen siehe Variablennamen in Kapitel 5 auf S. 83.

Unterdrücken der Anzeige mehrerer Daten-Sets

Wenn Sie bevorzugen, dass jeweils nur ein einziges Daten-Set zur Verfügung steht und Sie die Funktion für mehrere Daten-Sets deaktivieren möchten, gehen Sie wie folgt vor:

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Bearbeiten
  Optionen...

- Klicken Sie auf die Registerkarte Allgemein.

Aktivieren Sie Jeweils nur ein Daten-Set öffnen.

Für weitere Informationen siehe Optionen: Allgemein in Kapitel 45 auf S. 491.

Kapitel
7

Aufbereitung von Daten

Wenn Sie Daten im Daten-Editor eingegeben oder eine Datendatei geöffnet haben, können Sie ohne weitere Vorarbeit mit dem Erstellen von Berichten, Diagrammen und Analysen beginnen. Es gibt jedoch noch einige zusätzliche nützliche Funktionen zur Datenvorbereitung. Dazu gehören folgende Möglichkeiten:

- Zuweisen von Variableneigenschaften, welche die Daten beschreiben und festlegen, wie bestimmte Werte behandelt werden sollen.
- Identifizieren von Fällen die eventuell doppelte Informationen enthalten und Auschluss dieser Fälle aus den Analysen oder Löschen der Fälle aus der Datendatei.
- Erstellen neuer Variablen mit einigen verschiedenen Kategorien, die für Wertebereiche aus Variablen mit einer großen Anzahl möglicher Werte stehen.

Variableneigenschaften

Die im Daten-Editor in der Datenansicht eingegebenen Rohdaten oder die aus einem externen Dateiformat eingelesenen Daten (zum Beispiel aus einer Excel-Tabellenkalkulationsdatei einer Textdatei) verfügen noch nicht über einige spezielle, möglicherweise sehr nützliche Variableneigenschaften, wie zum Beispiel:

- Definition aussagekräftiger Wertelabels für numerische Codes (beispielsweise 0 = *Männlich* und 1 = *Weiblich*).
- Kennzeichnung fehlender Werte mit Codes (beispielsweise 99 = *Nicht zutreffend*).
- Zuweisung von Messniveaus (nominal, ordinal oder metrisch).

Diese und viele weitere Variableneigenschaften können in der Variablenansicht des Daten-Editors zugewiesen werden. Darüber hinaus stehen Ihnen zahlreiche Hilfsmittel zur Durchführung dieses Vorgangs zur Verfügung:

- **Variableneigenschaften definieren**. Diese Funktion unterstützt Sie bei der Definition von aussagekräftigen Wertelabels und fehlenden Werten. Sie ist besonders hilfreich bei der Aufbereitung kategorialer Daten mit numerischen Codes für Kategorienwerte.
- **Dateneigenschaften kopieren**. Mit dieser Funktion können Sie eine vorhandene Datendatei im SPSS-Format als Vorlage für die Datei- und Variableneigenschaften in der aktuellen Datendatei verwenden. Dies ist besonders nützlich, wenn Sie regelmäßig externe Datendateien mit ähnlichem Datenbestand verwenden (beispielsweise Monatsberichte im Excel-Format).

Definieren von Variableneigenschaften

Die Funktion “Variableneigenschaften definieren” ist daraufhin konzipiert, Ihnen das Erstellen von beschreibenden Wertelabels für kategoriale, nominale oder ordinale Variablen zu erleichtern. Mithilfe dieser Funktion können folgende Vorgänge durchgeführt werden:

- Durchsuchen der tatsächlichen Datenwerte und Auflisten aller eindeutigen Datenwerte für die ausgewählten Variablen.
- Ermitteln der Werte ohne Label und Bereitstellen einer Funktion zur automatischen Beschriftung.
- Kopieren definierter Wertelabels aus einer anderen Variablen in die ausgewählte Variable oder von der ausgewählten Variablen in mehrere zusätzliche Variablen.

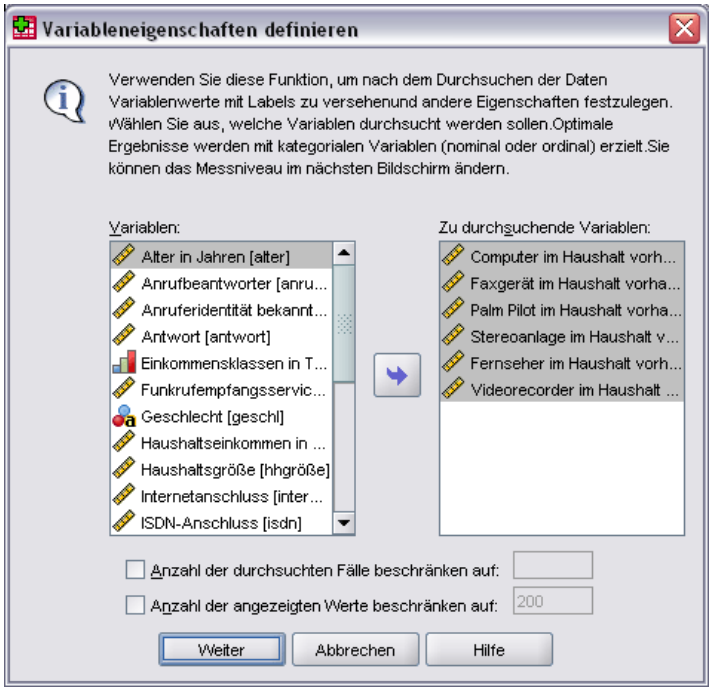
Anmerkung: Wenn von der Funktion “Variableneigenschaften definieren” keine Fälle durchsucht werden sollen, geben Sie “0” als Anzahl der zu durchsuchenden Fälle ein.

So definieren Sie Variableneigenschaften:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
Variableneigenschaften definieren...

Abbildung 7-1
Erstes Dialogfeld für die Auswahl zu definierender Variablen

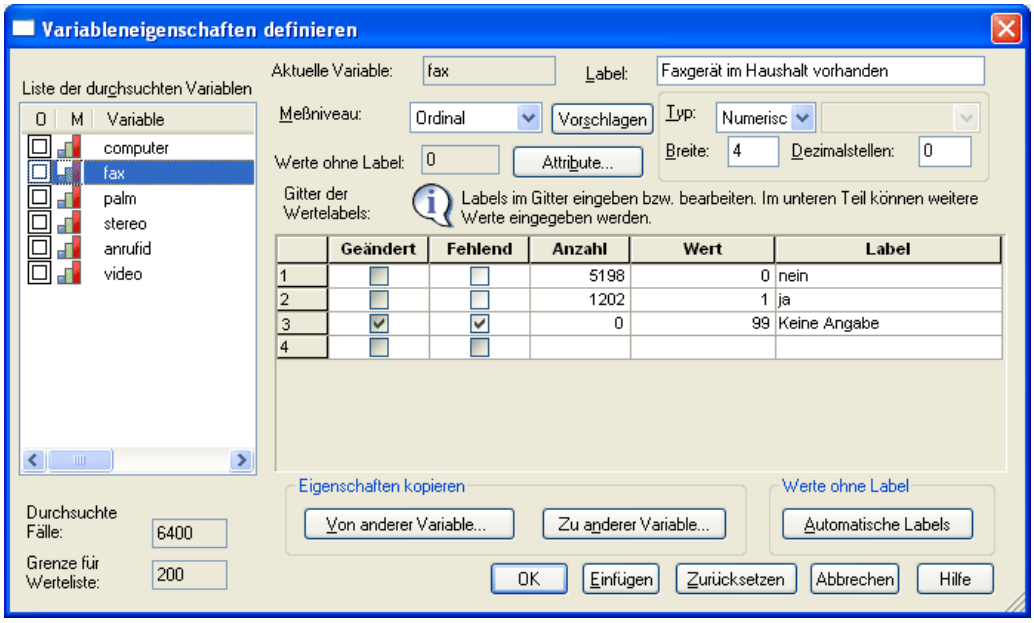


▶ Wählen Sie die numerischen oder String-Variablen aus, für die Wertelabels erstellt oder andere Variableneigenschaften wie fehlende Werte oder beschreibende Variablenlabels definiert bzw. geändert werden sollen.

- Geben Sie die Anzahl der Fälle an, die durchsucht und bei Erstellung einer Liste mit eindeutigen Werten berücksichtigt werden sollen. Dies ist insbesondere bei Datendateien mit einer großen Anzahl an Fällen nützlich, da sich das Durchsuchen der gesamten Datendatei in diesem Fall sehr zeitaufwendig gestalten würde.
- Geben Sie eine Obergrenze für die Anzahl der eindeutigen Werte an, die angezeigt werden soll. Durch diese Angabe wird in erster Linie vermieden, dass Hunderte, Tausende oder sogar Millionen von Werten für metrische Variablen bzw. stetige Intervall- oder Verhältnisvariablen aufgelistet werden.
- Klicken Sie auf Weiter, um das Hauptdialogfeld "Variableneigenschaften definieren" zu öffnen.
- Wählen Sie eine Variable aus, für die Wertelabels erstellt oder andere Variableneigenschaften definiert bzw. geändert werden sollen.
- Geben Sie den Beschriftungstext für alle Werte ohne Label ein, die im "Gitter der Wertelabels" angezeigt werden.
- Wenn Wertelabels für Werte erstellt werden sollen, die nicht angezeigt werden, können Sie die Werte in der Spalte *Werte* unter dem letzten durchsuchten Wert eingeben.
- Wiederholen Sie diesen Vorgang für jede aufgeführte Variable, für die Wertelabels erstellt werden sollen.
- Klicken Sie auf OK, um die Wertelabels und Variableneigenschaften zuzuweisen.

Definieren von Wertelabels und anderen Variableneigenschaften

Abbildung 7-2
Hauptdialogfeld "Variableneigenschaften definieren"



Das Hauptdialogfeld “Variableneigenschaften definieren” enthält folgende Informationen über die durchsuchten Variablen:

Liste der durchsuchten Variablen. Wenn eine durchsuchte Variable Werte ohne zugewiesene Wertelabels enthält, wird dies durch ein Häkchen in der Spalte *Ohne Label* (*O.*) angezeigt.

So sortieren Sie die Variablenliste, um alle Variablen mit Werten ohne Label am Anfang der Liste anzuzeigen:

▶ Klicken Sie in der Liste der durchsuchten Variablen auf die Spaltenüberschrift *Ohne Label*.

Sie können auch nach Variablennamen oder Messniveau sortieren, indem Sie in der Liste der durchsuchten Variablen auf die entsprechende Spaltenüberschrift klicken.

Gitter der Wertelabels

- **Label.** Zeigt alle bereits definierten Wertelabels an. In dieser Spalte können Sie Labels hinzufügen oder ändern.
- **Wert.** Zeigt für jede ausgewählte Variable die eindeutigen Werte an. Diese Liste mit eindeutigen Werten beruht auf der Anzahl der durchsuchten Fälle. Wenn beispielsweise nur die ersten 100 Fälle in der Datendatei durchsucht wurden, gibt die Liste nur die in diesen Fällen auftretenden eindeutigen Werte wieder. Wenn die Datendatei zuvor nach der Variablen sortiert wurde, der Wertelabels zugewiesen werden sollen, werden in der Liste möglicherweise weitaus weniger eindeutige Werte angezeigt, als tatsächlich in den Daten vorhanden sind.
- **Anzahl.** Die Häufigkeit, mit der jeder Wert in den durchsuchten Fällen auftritt.
- **Fehlend.** Werte, für die definiert wurde, dass sie fehlende Daten darstellen. Sie können die Zuweisung fehlender Werte für die Kategorie ändern, indem Sie auf das Kontrollkästchen klicken. Ein Häkchen zeigt an, dass die Kategorie als benutzerdefinierte fehlende Kategorie definiert ist. Wenn für eine Variable bereits ein Bereich von Werten als benutzerdefiniert fehlend definiert ist (z. B. 90-99), können Sie für diese Variable mithilfe der Funktion “Variableneigenschaften definieren” keine Kategorien für fehlende Werte hinzufügen oder ändern. Für Variablen mit Bereichen von fehlenden Werten können die Kategorien für fehlende Werte in der Variablenansicht des Daten-Editors geändert werden. Für weitere Informationen siehe Fehlende Werte in Kapitel 5 auf S. 88.
- **Geändert.** Zeigt an, dass ein Wertelabel hinzugefügt oder geändert wurde.

Anmerkung: Wenn Sie im ersten Dialogfeld als Anzahl der zu durchsuchenden Fälle “0” angegeben haben, ist das Gitter der Wertelabels mit Ausnahme der für die Variable bereits definierten Wertelabels und/oder Kategorien für fehlende Werte zu Beginn leer. Außerdem ist die Schaltfläche Vorschlagen für das Messniveau deaktiviert.

Messniveau. Wertelabels sind in erster Linie für kategoriale, d. h. nominale und ordinale Variablen, sinnvoll. Zudem werden kategoriale Variablen in einigen Prozeduren anders behandelt als metrische Variablen, sodass das Zuweisen des richtigen Messniveaus unter Umständen wichtig ist. In der Standardeinstellung wird allen neuen numerischen Variablen das metrische Messniveau zugewiesen. Daher werden möglicherweise auch viele kategoriale Variablen zunächst als metrisch angezeigt.

Wenn Sie sich nicht sicher sind, welches Messniveau einer Variablen zugewiesen werden soll, klicken Sie auf Vorschlagen.

Eigenschaften kopieren. Sie können Wertelabels und andere Variableneigenschaften aus einer anderen Variablen in die gerade ausgewählte Variable oder aus der gerade ausgewählten Variablen in eine oder mehr andere Variablen kopieren.

Werte ohne Label. Zum automatischen Erstellen von Beschriftungen für Werte ohne Labels klicken Sie auf Automatische Labels.

Variablenlabel und Anzeigeformat

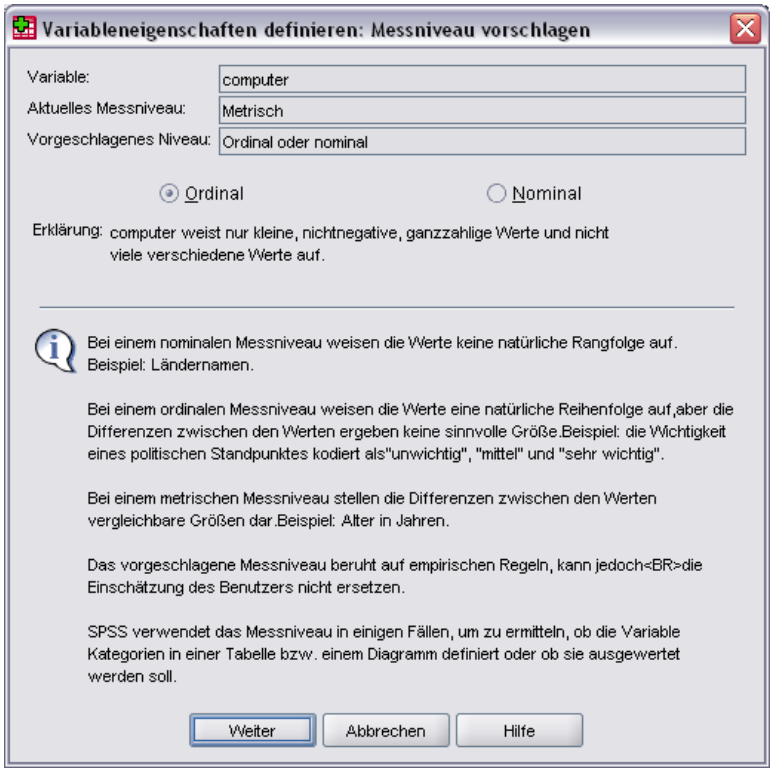
Sie können das beschreibende Variablenlabel und das Anzeigeformat ändern.

- Der grundlegende Typ der Variablen kann jedoch nicht geändert werden (String oder numerisch).
- Bei String-Variablen können Sie nur das Variablenlabel ändern, nicht jedoch das Anzeigeformat.
- Bei numerischen Variablen können Sie den numerischen Typ (z. B. numerisch, Datum, Dollar oder spezielle Währung), die Breite (Höchstzahl der Ziffern einschließlich aller Dezimal- oder Gruppentrennzeichen) sowie die Anzahl der Dezimalstellen ändern.
- Beim numerischen Datumsformat können Sie ein bestimmtes Datumsformat auswählen (z. B. tt-mm-jjjj, mm/tt/jj, jjjjttt).
- Bei benutzerdefinierten numerischen Formaten können Sie eine von fünf benutzerdefinierten Währungsformaten auswählen (CCA bis CCE). Für weitere Informationen siehe Optionen: Währung in Kapitel 45 auf S. 496.
- In der Spalte *Wert* wird ein Sternchen (*) angezeigt, wenn die angegebene Breite kleiner ist als die Breite der durchsuchten Werte oder der Werte für bereits vorhandene definierte Wertelabels bzw. für Kategorien für fehlende Werte.
- Ein Punkt (.) wird angezeigt, wenn die durchsuchten Werte oder die Werte für bereits vorhandene definierte Wertelabels bzw. Kategorien für fehlende Werte für den ausgewählten Anzeigetyp ungültig sind. Ein interner numerischer Wert von weniger als 86.400 ist beispielsweise für eine Variable im Datumsformat ungültig.

Zuweisen des Messniveaus

Wenn Sie im Hauptdialogfeld “Variableneigenschaften definieren” für das Messniveau auf Vorschlagen klicken, wird die aktuelle Variable anhand der durchsuchten Fälle und der definierten Wertelabels bewertet. Anschließend wird das Dialogfeld “Messniveau vorschlagen” mit einem Vorschlag für das Messniveau eingeblendet. Im Bereich “Erklärung” finden Sie eine kurze Beschreibung der Kriterien, nach denen das vorgeschlagene Messniveau ausgewählt wurde.

Abbildung 7-3
Dialogfeld "Messniveau vorschlagen"



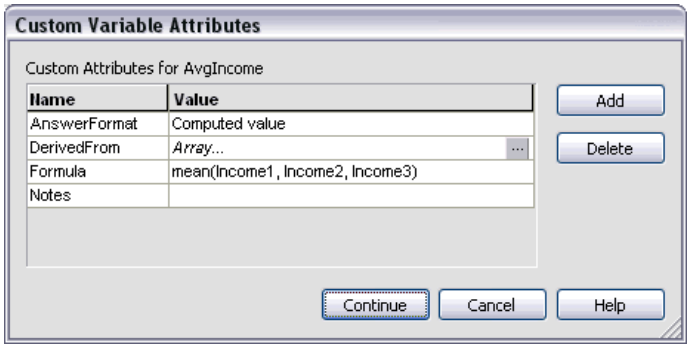
Anmerkung: Werte, die als fehlende Werte definiert wurden, werden bei der Berechnung des Messniveaus nicht berücksichtigt. In der Erklärung für das vorgeschlagene Messniveau kann beispielsweise darauf hingewiesen werden, dass der Vorschlag teilweise darauf beruht, dass die Variable keine negativen Werte enthält, obgleich sie in Wirklichkeit möglicherweise negative Werte enthält, die jedoch bereits als fehlende Werte definiert sind.

- Klicken Sie auf Weiter, um das vorgeschlagene Messniveau zu übernehmen, oder auf Abbrechen, um das Messniveau unverändert beizubehalten.

Benutzerdefinierte Variablenattribute

Mit der Schaltfläche Attribute unter "Variableneigenschaften definieren" wird das Dialogfeld "Benutzerdefinierte Variablenattribute" geöffnet. Neben den standardmäßigen Variablenattributen (z. B. Wertelabels, fehlende Werte, Messniveau) können Sie Ihre eigenen benutzerdefinierten Variablenattribute erstellen. Wie die standardmäßigen Variablenattribute werden auch die benutzerdefinierten Attribute zusammen mit Datendateien im SPSS-Format gespeichert.

Abbildung 7-4
Benutzerdefinierte Variablenattribute



Name. Für Attributnamen gelten dieselben Regeln wie für Variablennamen. Für weitere Informationen siehe Variablennamen in Kapitel 5 auf S. 83.

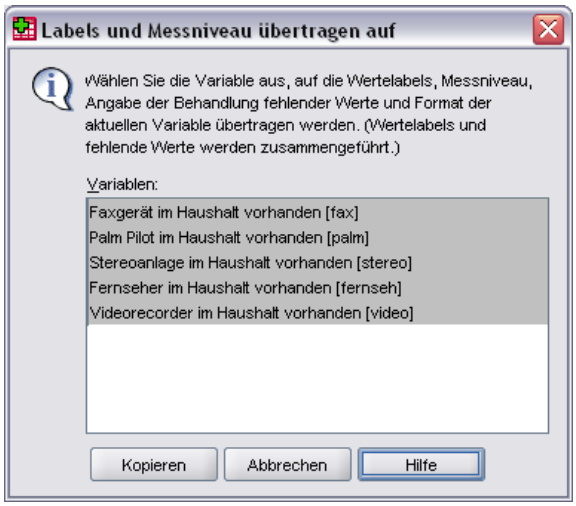
Wert. Der dem Attribut für die ausgewählte Variable zugewiesene Wert.

- Attributnamen, die mit einem Dollarzeichen beginnen, sind reserviert und können nicht geändert werden.
- Wenn der Text *Array...* in einer Wertezelle angezeigt wird, bedeutet dies, dass es sich dabei um ein **Attribut-Array** handelt, ein Attribut, das mehrere Werte enthält. Klicken Sie auf die Schaltfläche in der Zelle, um die Liste der Werte anzuzeigen.

Kopieren von Variableneigenschaften

Das Dialogfeld "Labels und Messniveau übertragen" wird angezeigt, wenn Sie im Hauptdialogfeld "Variableneigenschaften definieren" auf Von anderer Variable oder Zu anderer Variable klicken. In diesem Dialogfeld werden alle durchsuchten Variablen angezeigt, die mit dem Typ der aktuellen Variablen übereinstimmen (numerisch oder String). Bei String-Variablen muss auch die definierte Länge übereinstimmen.

Abbildung 7-5
Dialogfeld "Labels und Messniveau übertragen"



- Wählen Sie eine Variable aus, von der Wertelabels und andere Variableneigenschaften (außer dem Variablenlabel) kopiert werden sollen.

  oder

- Wählen Sie eine oder mehrere Variablen aus, in die Wertelabels und andere Variableneigenschaften kopiert werden sollen.

- Klicken Sie auf Kopieren, um die Wertelabels und das Messniveau zu kopieren.
  - Wertelabels und Kategorien für fehlende Werte, die bereits in der Zielvariablen vorhanden sind, werden nicht ersetzt.
  - Wertelabels und Kategorien für fehlende Werte, die noch nicht für die Zielvariable(n) definiert sind, werden den Wertelabels und den Kategorien für fehlende Werte der Zielvariablen hinzugefügt.
  - Das Messniveau für die Zielvariable(n) wird immer ersetzt.
  - Wenn entweder die Quell- oder die Zielvariable einen definierten Bereich von fehlenden Werten aufweist, werden die Definitionen für die fehlenden Werte nicht kopiert.

Mehrfachantworten-Sets

In benutzerdefinierten Tabellen und der Diagrammerstellung kann eine besondere Art von "Variable" verwendet werden, die als **Mehrfachantworten-Set** bezeichnet wird. Bei Mehrfachantworten-Sets handelt es sich nicht um "Variablen" im üblichen Sinn. Mehrfachantworten-Sets können nicht im Daten-Editor angezeigt werden, sie werden von anderen Prozeduren nicht erkannt. Mehrfachantworten-Sets verwenden mehrere Variablen, um Antworten auf Fragen aufzuzeichnen, auf welche der Befragte mehr als eine Antwort geben kann. Sie werden wie kategoriale Variablen behandelt und bieten weitestgehend dieselben Möglichkeiten wie kategoriale Variablen.

Mehrfachantworten-Sets werden aus mehreren Variablen in der Datendatei gebildet. Beim Mehrfachantworten-Set handelt es sich um ein spezielles Konstrukt innerhalb einer Datendatei im SPSS-Format. Mehrfachantworten-Sets können in einer Datendatei im SPSS-Format definiert und gespeichert werden. Das Importieren oder Exportieren von Mehrfachantworten-Sets in bzw. aus anderen Dateiformaten ist jedoch nicht möglich. (Mit der Option "Dateneigenschaften kopieren" können Sie Mehrfachantworten-Sets aus anderen SPSS-Datendateien kopieren. Sie finden diese Option im Menü "Daten" im Fenster "Daten-Editor".Für weitere Informationen siehe Kopieren von Dateneigenschaften auf S. 120.)

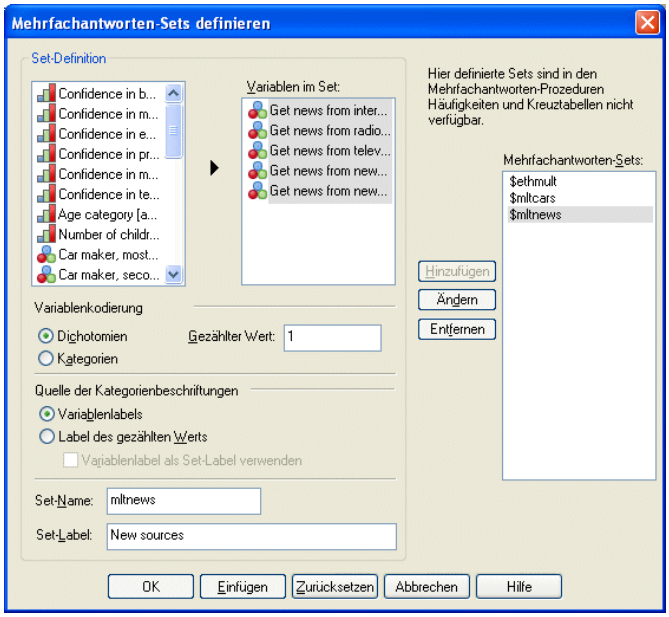
Definieren von Mehrfachantworten-Sets

So definieren Sie Mehrfachantworten-Sets:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Daten
  Mehrfachantworten-Sets definieren...

Abbildung 7-6
Dialogfeld "Mehrfachantworten-Sets definieren"



- Wählen Sie mindestens zwei Variablen aus. Wenn Ihre Variablen als Dichotomien kodiert sind, geben Sie an, welcher Wert gezählt werden soll.
- Geben Sie einen eindeutigen Namen für jedes Mehrfachantworten-Set ein. Der Name darf maximal 63 Byte lang sein. Dem Set-Namen wird automatisch ein Dollarzeichen vorangestellt.
- Geben Sie einen aussagekräftigen Namen für das Set ein. (Optional.)
- Klicken Sie auf Hinzufügen, um das Mehrfachantworten-Set zur Liste der definierten Sets hinzuzufügen.

Dichotomien

Ein Set aus dichotomen Variablen enthält gewöhnlich mehrere dichotome Variablen, d. h. Variablen mit nur zwei möglichen Werten der folgenden Art: ja/nein, vorhanden/nicht vorhanden oder angekreuzt/nicht angekreuzt. Wenngleich die Variablen nicht immer streng dichotom sein müssen, werden alle Variablen des Sets auf dieselbe Weise kodiert, und für "Gezählter Wert" gilt die Bedingung "positiv/vorhanden/angekreuzt".

In einer Umfrage wird beispielsweise die folgende Frage gestellt: "Aus welchen der folgenden Quellen beziehen Sie Nachrichten?", wobei fünf Antwortmöglichkeiten zur Auswahl stehen. Die Befragten können mehrere Antworten geben, indem sie mehrere Felder neben den einzelnen Antworten ankreuzen. Die fünf Antworten entsprechen fünf Variablen in der Datendatei, wobei 0 für *Nein* (nicht angekreuzt) und 1 für *Ja* (angekreuzt) steht. Im Set aus dichotomen Variablen ist der "Gezählte Wert" gleich 1.

Die Beispieldatendatei *survey_sample.sav* verfügt bereits über drei definierte Mehrfachantworten-Sets. *$mltnews* ist ein Set aus dichotomen Variablen.

- Wählen Sie *$mltnews* in der Liste für Mehrfachantworten-Sets aus, indem Sie darauf klicken.

Daraufhin werden die Variablen und Einstellungen angezeigt, die für die Definition dieses Mehrfachantworten-Sets verwendet wurden.

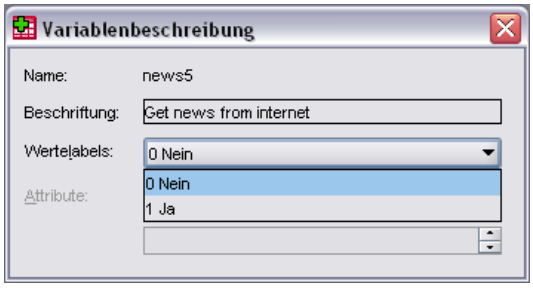
- Die Liste “Variablen im Set” zeigt die fünf Variablen an, mit denen das Mehrfachantworten-Set erstellt wurde.
- Im Gruppenfeld für die Variablenkodierung wird angezeigt, dass es sich bei den Variablen um dichotome Variablen handelt.
- Für “Gezählter Wert” wird 1 angegeben.

▶ Wählen Sie eine Variable in der Liste “Variablen im Set” aus, indem Sie darauf klicken.

▶ Klicken Sie mit der rechten Maustaste auf die Variable und wählen Sie im Kontextmenü die Option Variablenbeschreibung aus.

▶ Klicken Sie im Fenster “Variablenbeschreibung” auf den Abwärtspfeil neben der Dropdown-Liste “Wertelabels”, um die Liste mit den definierten Wertelabels vollständig anzuzeigen.

Abbildung 7-7
Variablenbeschreibung für eine Quellvariable für ein Set aus dichotomen Variablen



Die Wertelabels geben an, dass es sich bei der Variablen um eine dichotome Variable mit den Werten 0 und 1 handelt, die jeweils für *Nein* und *Ja* stehen. Alle fünf Variablen in der Liste sind auf dieselbe Weise kodiert, und der Wert 1 (Code für *Ja*) bildet den gezählten Wert für das Set aus dichotomen Variablen.

Kategorien

Ein Set aus kategorialen Variablen besteht aus Mehrfachvariablen, die alle auf dieselbe Weise kodiert wurden, häufig mit zahlreichen möglichen Antwortkategorien. Ein Umfragethema lautet beispielsweise “Nennen Sie bis zu drei Nationalitäten, die am besten Ihre ethnische Herkunft beschreiben”. Zu diesem Thema gibt es Hunderte von möglichen Antworten. Für die Kodierung wird die Liste jedoch auf die 40 häufigsten Nationalitäten begrenzt und alle anderen auf die Kategorie “Andere” verwiesen. In der Datendatei werden die Auswahlmöglichkeiten zu drei Variablen, wobei jede über 41 Kategorien verfügt (40 kodierte Nationalitäten und eine Kategorie “Andere”).

In der Beispieldatendatei bilden *$ethmult* und *$mltcars* Sets aus kategorialen Variablen.

Quelle der Kategorienbeschriftungen

Bei mehreren Dichotomien können Sie festlegen, wie die Sets beschriftet werden.

- **Variablenlabels.** Hierbei werden die definierten Variablenlabels (oder Variablennamen für Variablen ohne definierte Variablenlabels) als Beschriftungen für die Kategorien des Sets verwendet. Wenn zum Beispiel alle Variablen im Set dasselbe Wertelabel (oder keine Wertelabels) für den gezählten Wert aufweisen (z. B. *Ja*), sollten Sie die Variablenlabels als Beschriftungen für die Kategorien des Sets verwenden.
- **Label des gezählten Werts.** Hierbei werden die definierten Wertelabels der gezählten Werte als Beschriftungen für die Kategorien des Sets verwendet. Wählen Sie diese Option nur aus, wenn alle Variablen ein definiertes Wertelabel für den gezählten Wert aufweisen und sich das Wertelabel für den gezählten Wert in jeder Variable unterscheidet.
- **Variablenlabel als Set-Label verwenden.** Wenn Sie Label des gezählten Werts auswählen, können Sie auch das Variablenlabel der ersten Variablen im Set mit einem definierten Variablenlabel als Set-Label verwenden. Wenn keine der Variablen im Set definierte ein definiertes Variablenlabel aufweist, wird der Name der ersten Variable im Set als Set-Label verwendet.

Kopieren von Dateneigenschaften

Mit dem Assistenten zum Kopieren von Dateneigenschaften können Sie eine externe SPSS-Datendatei als Vorlage für die Definition von Datei- und Variableneigenschaften in der Arbeitsdatei verwenden. Sie können außerdem Variablen in der Arbeitsdatei als Vorlagen für andere Variablen in der Arbeitsdatei verwenden. Sie verfügen über folgende Möglichkeiten:

- Kopieren ausgewählter Dateieigenschaften aus einer externen Datendatei oder einem geöffneten Daten-Set in die Arbeitsdatei. Zu den Dateieigenschaften gehören Dokumente, Dateilabel, Mehrfachantworten-Sets, Variablen-Sets und Gewichtung.
- Kopieren ausgewählter Variableneigenschaften aus einer externen Datendatei oder einem geöffneten Daten-Set in entsprechende Variablen in der Arbeitsdatei. Zu den Variableneigenschaften gehören Wertelabels, fehlende Werte, Messniveau, Variablenlabels, Druck- und Schreibformate, Ausrichtung und Spaltenbreite (im Daten-Editor).
- Kopieren von ausgewählten Variableneigenschaften einer Variablen in einer externen Datendatei, einem geöffneten Daten-Set oder der Arbeitsdatei in viele Variablen in der Arbeitsdatei.
- Erstellen neuer Variablen in der Arbeitsdatei anhand von ausgewählten Variablen in einer externen Datendatei oder einem geöffneten Daten-Set.

Beim Kopieren von Dateneigenschaften gelten die folgenden allgemeinen Regeln:

- Wenn eine externe Datendatei als Quelldatendatei verwendet werden soll, muss diese Datendatei SPSS-Format aufweisen.
- Wenn die Arbeitsdatei als Quelldatendatei verwendet wird, muss die Datei mindestens eine Variable enthalten. Vollständig leere Arbeitsdateien können nicht als Quelldatendateien verwendet werden.

- Bereits definierte Eigenschaften in der Arbeitsdatei können nicht von undefinierten (leeren) Eigenschaften aus dem Quelldaten-Set überschrieben werden.
- Variableneigenschaften aus Quellvariablen können nur in Zielvariablen mit folgendem Datentyp kopiert werden: String (alphanumerisch) oder numerisch (einschließlich numerisch, Datum und Währung).

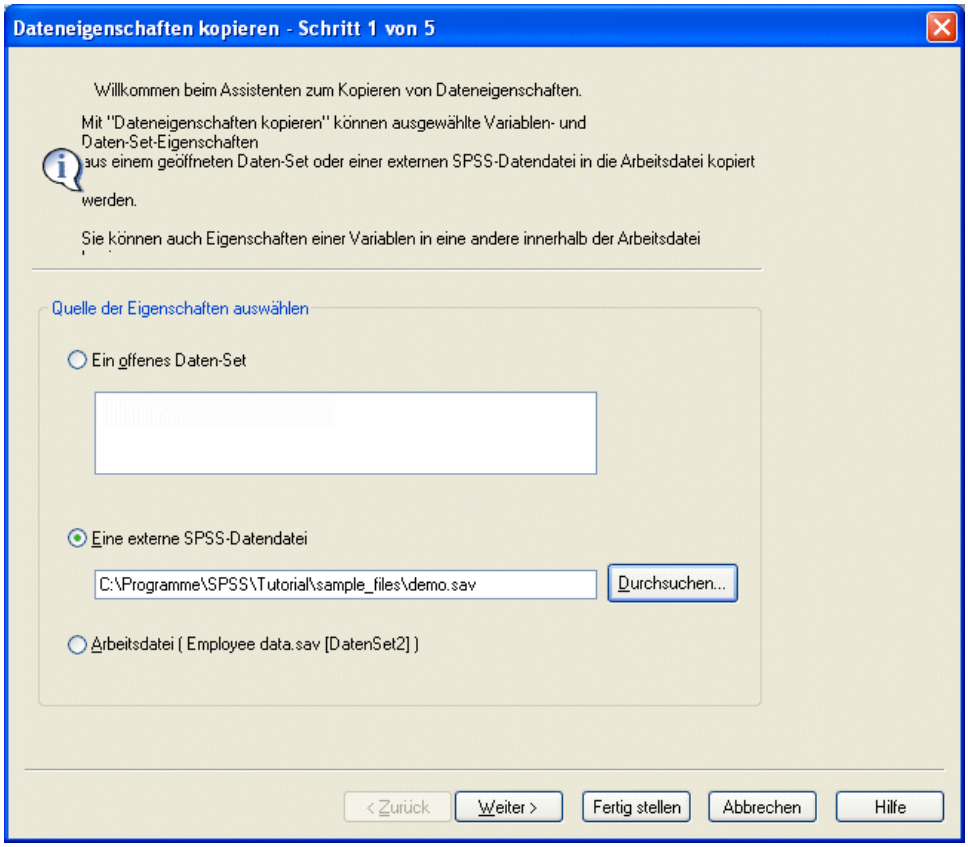
Anmerkung: Die Option “Dateneigenschaften kopieren” ersetzt die Option “Datenlexikon zuweisen”, die in der Vorversion im Menü “Datei” verfügbar war.

So kopieren Sie Dateneigenschaften:

▶ Wählen Sie die folgenden Befehle aus den Menüs im Fenster “Daten-Editor” aus:

Daten
  Dateneigenschaften kopieren...

Abbildung 7-8
Assistent zum Kopieren von Dateneigenschaften: Schritt 1



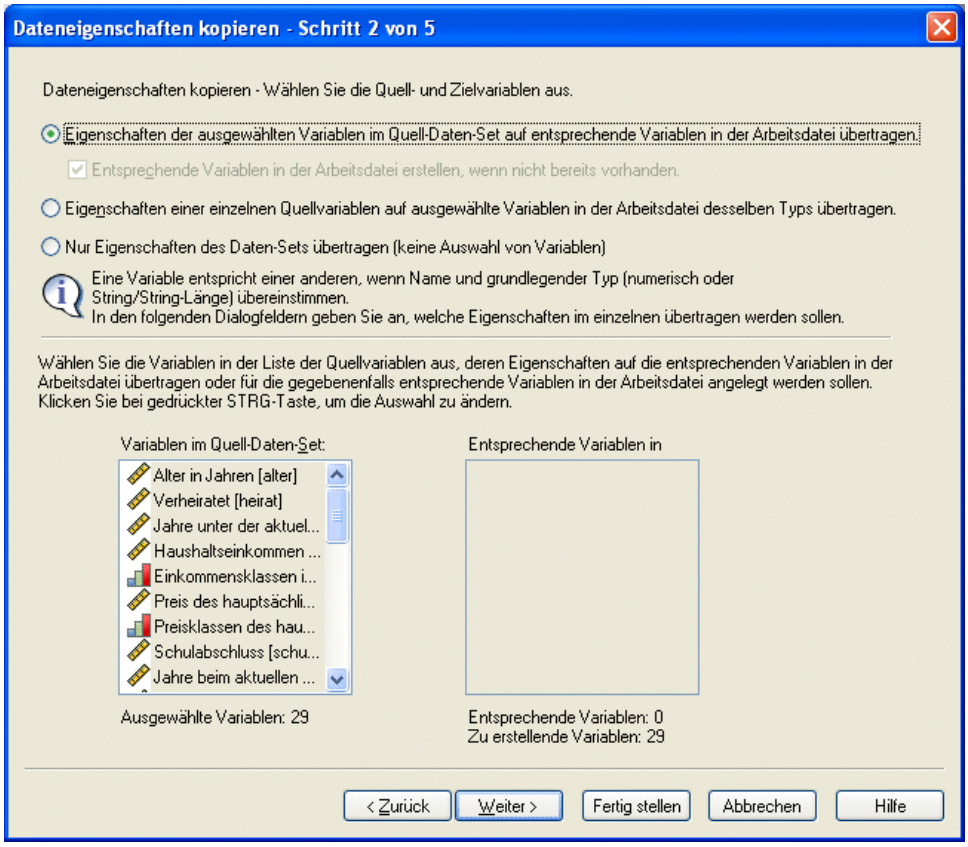
▶ Wählen Sie die Datendatei mit den zu kopierenden Datei- und/oder Variableneigenschaften aus. Dabei kann es sich um ein gegenwärtig geöffnetes Daten-Set, eine externe Datendatei im SPSS-Format oder die Arbeitsdatei handeln.

▶ Befolgen Sie die Anweisungen im Assistenten zum Kopieren von Dateneigenschaften Schritt für Schritt.

Auswählen von Quell- und Zielvariablen

In diesem Schritt können Sie die Quellvariablen mit den zu kopierenden Variableneigenschaften und die Zielvariablen angeben, die diese Variableneigenschaften erhalten sollen.

Abbildung 7-9
Assistent zum Kopieren von Dateneigenschaften: Schritt 2



Eigenschaften der ausgewählten Variablen in der Quelldatei auf entsprechende Variablen in der Arbeitsdatei übertragen. Die Variableneigenschaften werden aus einer oder mehreren ausgewählten Quellvariablen in entsprechende Variablen in der Arbeitsdatei kopiert. Variablen "stimmen überein", wenn sowohl der Variablenname als auch der Variablentyp (String oder numerisch) identisch sind. Bei String-Variablen muss auch die definierte Länge übereinstimmen. In der Standardeinstellung werden in den beiden Variablenlisten nur übereinstimmende Variablen angezeigt.

- **Entsprechende Variablen in der Arbeitsdatei erstellen, wenn nicht bereits vorhanden.** Dadurch wird die Liste mit Quellvariablen so aktualisiert, dass alle Variablen der Quelldatendatei angezeigt werden. Wenn Sie Quellvariablen auswählen, die (dem Variablennamen nach) nicht in der Arbeitsdatei vorhanden sind, werden in der Arbeitsdatei neue Variablen mit den Variablennamen und -eigenschaften aus der Quelldatendatei erstellt.

Wenn die Arbeitsdatei keine Variablen enthält (ein leeres, neues Daten-Set), werden alle Variablen in der Quelldatendatei angezeigt und in der Arbeitsdatei werden automatisch neue Variablen anhand der ausgewählten Quellvariablen erstellt.

Eigenschaften einer einzelnen Quellvariablen auf ausgewählte Variablen in der Arbeitsdatei desselben Typs übertragen. Die Variableneigenschaften einer in der Liste mit Quellvariablen ausgewählten Variablen können auf mehrere in der Liste “Variablen in Arbeitsdatei” ausgewählte Variablen übertragen werden. In der Liste “Variablen in Arbeitsdatei” werden nur Variablen angezeigt, deren Typ mit der in der Liste “Variablen in Quelldatei” ausgewählten Variablen übereinstimmt (numerisch oder String). Bei einer String-Variablen werden nur Strings angezeigt, deren definierte Länge mit der aus der Quellvariablen übereinstimmt. Diese Option ist nicht verfügbar, wenn die Arbeitsdatei keine Variablen enthält.

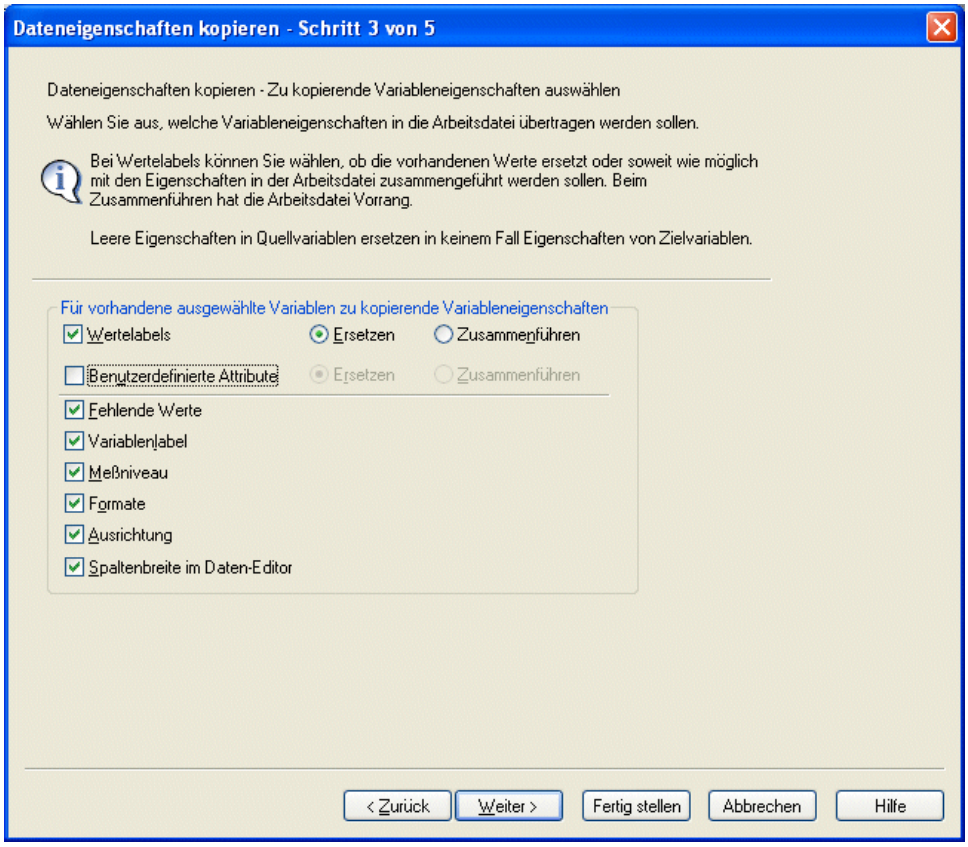
Anmerkung: Mit dieser Option können Sie keine neuen Variablen in der Arbeitsdatei erstellen.

Nur Eigenschaften des Daten-Sets übertragen, keine Auswahl von Variablen. In die Arbeitsdatei werden nur die Dateieigenschaften übertragen (z. B. Dokumente, Dateilabel, Gewichtung). Es werden keine Variableneigenschaften übertragen. Diese Option ist nicht verfügbar, wenn die Arbeitsdatei gleichzeitig die Quelldatendatei darstellt.

Auswählen von Variableneigenschaften zum Kopieren

Sie können ausgewählte Variableneigenschaften aus den Quellvariablen in die Zielvariablen kopieren. Bereits definierte Eigenschaften in den Zielvariablen können nicht von undefinierten (leeren) Eigenschaften aus den Quellvariablen überschrieben werden.

Abbildung 7-10
Assistent zum Kopieren von Dateneigenschaften: Schritt 3



Wertelabels. Wertelabels sind umschreibende Beschriftungen, die Datenwerten zugeordnet sind. Wertelabels werden häufig verwendet, wenn numerische Datenwerte zur Darstellung nichtnumerischer Kategorien verwendet werden (z. B. die Codes 1 und 2 für *Männlich* und *Weiblich*). Sie können Wertelabels in den Zielvariablen ersetzen oder zusammenführen.

- Durch Ersetzen werden alle definierten Wertelabels der Zielvariablen gelöscht und durch die definierten Wertelabels aus der Quellvariablen ersetzt.
- Durch Zusammenführen werden die definierten Wertelabels der Quellvariablen mit allen definierten Wertelabels der Zielvariablen zusammengeführt. Wenn für denselben Wert sowohl in den Quell- als auch in den Zielvariablen ein Wertelabel definiert ist, wird das Wertelabel in der Zielvariablen nicht geändert.

Benutzerdefinierte Attribute. Benutzerdefinierte Variablenattribute. Für weitere Informationen siehe Benutzerdefinierte Variablenattribute in Kapitel 5 auf S. 91.

- Durch Ersetzen werden alle benutzderdefinierten Attribute der Zielvariablen gelöscht und durch die definierten Attribute aus der Quellvariablen ersetzt.
- Durch Zusammenführen werden die definierten Attribute der Quellvariablen mit allen definierten Attributen der Zielvariablen zusammengeführt.

Fehlende Werte. Fehlende Werte sind Werte, die anstelle von fehlenden Werten eingesetzt werden (z. B. 98 für *Weiß nicht* und 99 für *Nicht zutreffend*). In der Regel verfügen auch diese Werte über definierte Wertelabels. Sie beschreiben, welche Bedeutung die Codes für die fehlenden Werte tragen. Alle für die Zielvariable definierten fehlenden Werte werden gelöscht und durch die entsprechenden fehlenden Werte aus der Quellvariablen ersetzt.

Variablenlabel. Aussagekräftige Variablenlabels können Leerzeichen und reservierte Zeichen enthalten, die in Variablennamen nicht zulässig sind. Beim Kopieren von Variableneigenschaften aus einer Quellvariablen in mehrere Zielvariablen sollten Sie sich die Verwendung dieser Option genau überlegen.

Messniveau. Das Messniveau kann nominal, ordinal oder metrisch sein. Bei Prozeduren, die zwischen verschiedenen Messniveaus unterscheiden, werden sowohl nominale als auch ordinale Messniveaus als **kategorial** betrachtet.

Formate. Bei numerischen Variablen wird über die Formatangabe der numerische Typ (z. B. numerisch, Datum oder Währung), die Breite (Gesamtzahl der angezeigten Zeichen einschließlich der führenden Zeichen, der Abschlusszeichen und des Dezimaltrennzeichens) sowie die Anzahl der angezeigten Dezimalstellen festgelegt. Diese Option gilt nicht für String-Variablen.

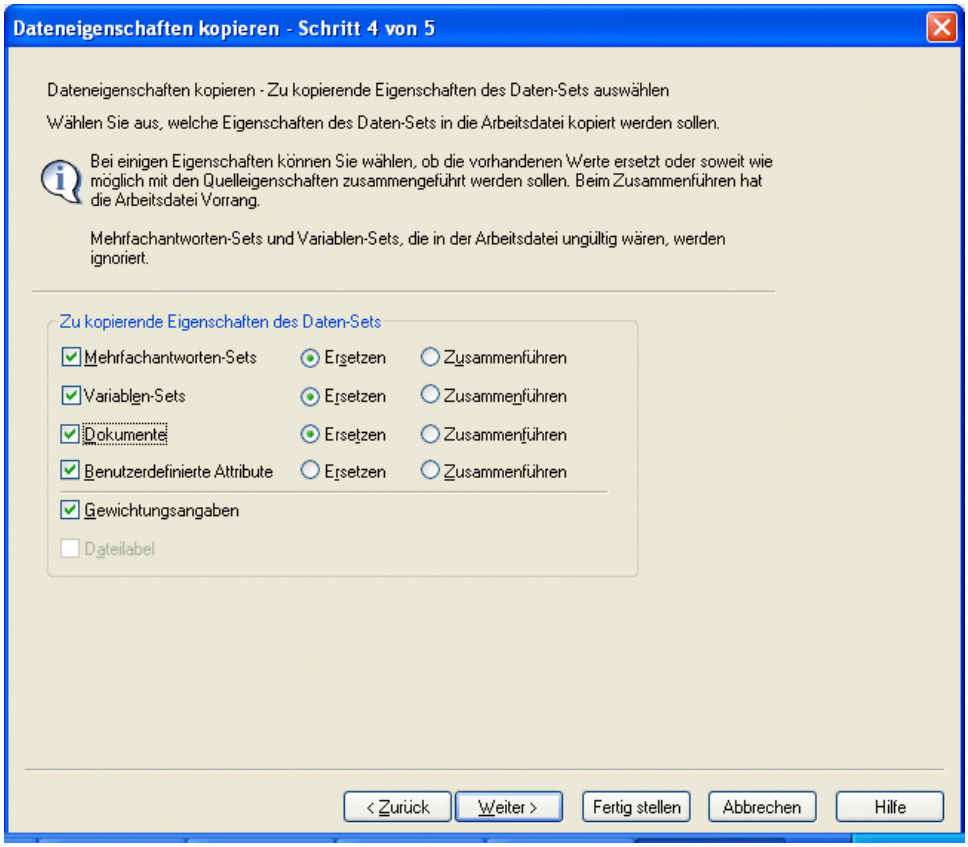
Ausrichtung. Dies betrifft nur die Ausrichtung der Daten in der Datenansicht des Daten-Editors (linksbündig, rechtsbündig, zentriert).

Spaltenbreite im Daten-Editor. Dies betrifft nur die Spaltenbreite in der Datenansicht des Daten-Editors.

Kopieren der (Datei-)Eigenschaften eines Daten-Sets

In einer Quelldatendatei können die globalen Eigenschaften der Daten-Sets ausgewählt und in die Arbeitsdatei übertragen werden. (Diese Option ist nicht verfügbar, wenn die Arbeitsdatei gleichzeitig die Quelldatendatei ist.)

Abbildung 7-11
Assistent zum Kopieren von Dateneigenschaften: Schritt 4



Mehrfachantworten-Sets. Überträgt Definitionen des Mehrfachantworten-Sets aus der Quelldatendatei in die Arbeitsdatei. (*Anmerkung*: Mehrfachantworten-Sets werden gegenwärtig nur in der Diagrammerstellung und dem Erweiterungsmodul "Tables" (Tabellen) verwendet.)

- Mehrfachantworten-Sets in der Quelldatendatei, die Variablen ohne Entsprechungen in der Arbeitsdatei enthalten, werden solange ignoriert, bis die entsprechenden Zielvariablen anhand der in Schritt 2 (Auswählen von Quell- und Zielvariablen) angegebenen Einstellungen im Assistenten zum Kopieren von Dateneigenschaften erstellt worden sind.
- Durch Ersetzen werden alle Mehrfachantworten-Sets in der Arbeitsdatei gelöscht und durch die Mehrfachantworten-Sets aus der Quelldatendatei ersetzt.
- Durch Zusammenführen werden den gesamten Mehrfachantworten-Sets in der Arbeitsdatei die Mehrfachantworten-Sets aus der Quelldatendatei hinzugefügt. Wenn beide Dateien ein Set mit demselben Namen enthalten, bleibt das Set in der Arbeitsdatei unverändert.

Variablen-Sets. Mit Variablen-Sets werden die Variablenlisten festgelegt, die in den Dialogfeldern angezeigt werden sollen. Die Definition der Variablen-Sets erfolgt im Menü "Extras" unter "Sets definieren".

- Sets in der Quelldatendatei, die Variablen ohne Entsprechungen in der Arbeitsdatei enthalten, werden solange ignoriert, bis die entsprechenden Zielvariablen anhand der in Schritt 2 (Auswählen von Quell- und Zielvariablen) angegebenen Einstellungen im Assistenten zum Kopieren von Dateneigenschaften erstellt worden sind.
- Durch Ersetzen werden alle Variablen-Sets in der Arbeitsdatei gelöscht und durch Variablen-Sets aus der Quelldatendatei ersetzt.
- Durch Zusammenführen werden den gesamten Variablen-Sets in der Arbeitsdatei die Variablen-Sets aus der Quelldatendatei hinzugefügt. Wenn beide Dateien ein Set mit demselben Namen enthalten, bleibt das Set in der Arbeitsdatei unverändert.

Dokumente. Anmerkungen, die mit dem Befehl `DOCUMENT` an die Datendatei angefügt sind.

- Durch Ersetzen werden alle Dokumente in der Arbeitsdatei gelöscht und durch Dokumente aus der Quelldatendatei ersetzt.
- Durch Zusammenführen werden die Dokumente aus der Quelldatendatei und der Arbeitsdatei kombiniert. Quelldokumente, die in der Arbeitsdatei nicht vorhanden sind, werden der Arbeitsdatei hinzugefügt. Anschließend werden alle Dokumente nach Datum sortiert.

Benutzerdefinierte Attribute. Benutzerdefinierte Datendatei-Attribute, die in der Regel mit dem Befehl `DATAFILE ATTRIBUTE` der Befehlssyntax erstellt werden.

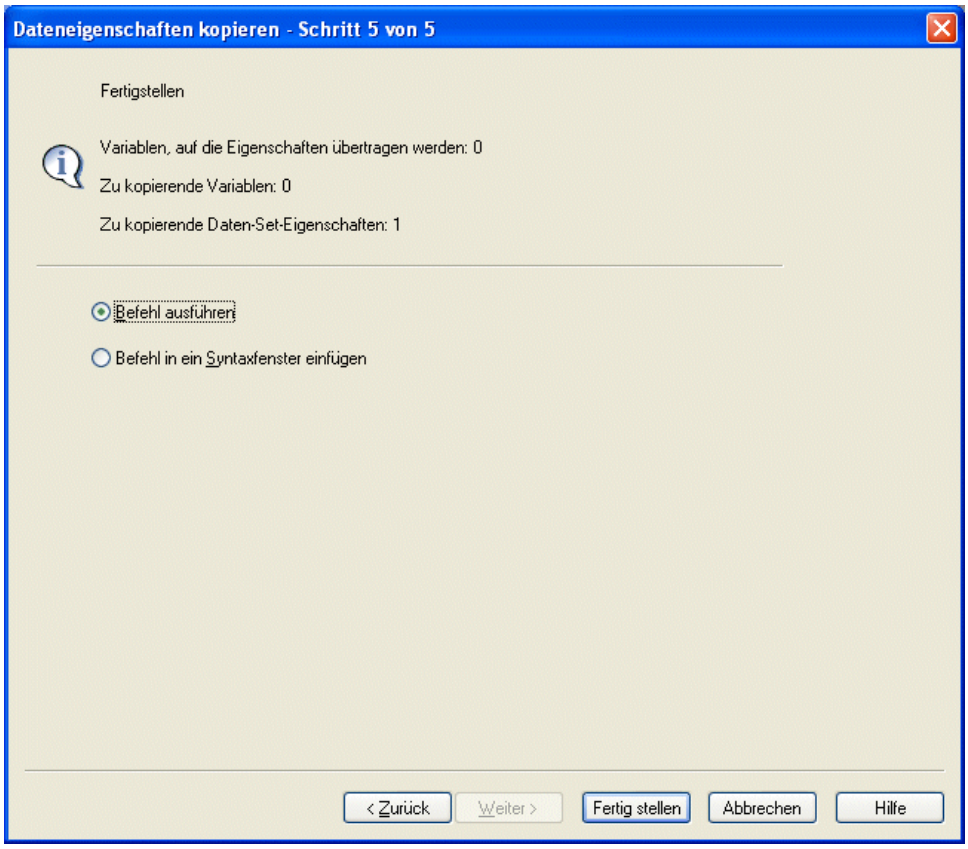
- Durch Ersetzen werden alle vorhandenen benutzerdefinierten Datendatei-Attribute in der Arbeitsdatei gelöscht und durch Datendatei-Attribute aus der Quelldatendatei ersetzt.
- Durch Zusammenführen werden die Datendatei-Attribute aus der Quelldatendatei und der Arbeitsdatei kombiniert. Eindeutige Attributenamen in der Quelldatei, die in der Arbeitsdatei nicht vorhanden sind, werden der Arbeitsdatei hinzugefügt. Wenn derselbe Attributname in beiden Datendateien vorhanden ist, bleibt der Attributname in der Arbeitsdatei unverändert.

Gewichtungsangaben. Mit dieser Option werden Fälle mit der aktuellen Gewichtungsvariablen der Quelldatei gewichtet, sofern in der Arbeitsdatei eine entsprechende Variable vorhanden ist. Dadurch werden alle Gewichtungen überschrieben, die bis dahin in der Arbeitsdatei gültig waren.

Dateilabel. Eine Beschriftung, die einer Datendatei über den Befehl `FILE LABEL` zugeordnet ist.

Ergebnisse

Abbildung 7-12
Assistent zum Kopieren von Dateneigenschaften: Schritt 5



Im letzten Schritt des Assistenten zum Kopieren von Dateieigenschaften erhalten Sie Informationen über die Anzahl der Variablen, für die Variableneigenschaften aus der Quelldatendatei kopiert werden sollen, sowie Informationen über die Anzahl der neu zu erstellenden Variablen und über die Anzahl der Eigenschaften, die aus den Daten-Sets kopiert werden sollen.

Sie können die erstellte Befehlssyntax auch in ein Syntax-Fenster einfügen und sie zur späteren Verwendung speichern.

Ermitteln doppelter Fälle

"Doppelte" Fälle können aus einer Vielzahl von Gründen in Ihren Daten vorkommen. Dazu gehören:

- Dateneingabefehler, bei denen derselbe Fall versehentlich mehrmals eingegeben wurde.

- Mehrere Fälle haben denselben Primär-ID-Wert, aber verschiedene Sekundär-ID-Werte, beispielsweise bei Familienmitgliedern, die alle im selben Haus leben.
- Mehrere Fälle stellen denselben Fall dar, jedoch mit unterschiedlichen Werten für die Variablen, die nicht zur Identifizierung des Falles dienen, beispielsweise mehrere Kaufvorgänge, die von derselben Person oder demselben Unternehmen für verschiedene Produkte oder zu verschiedenen Zeitpunkten durchgeführt wurden.

Mit “Doppelte Fälle ermitteln” haben Sie bei der Definition von *doppelt* sehr große Freiheiten und gewisse Steuerungsmöglichkeiten bei der automatischen Unterscheidung von primären Fällen und doppelten Fällen.

So können Sie doppelte Fälle ermitteln und markieren:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Doppelte Fälle ermitteln...

▶ Wählen Sie eine oder mehrere Variablen für die Identifikation übereinstimmender Fälle aus.

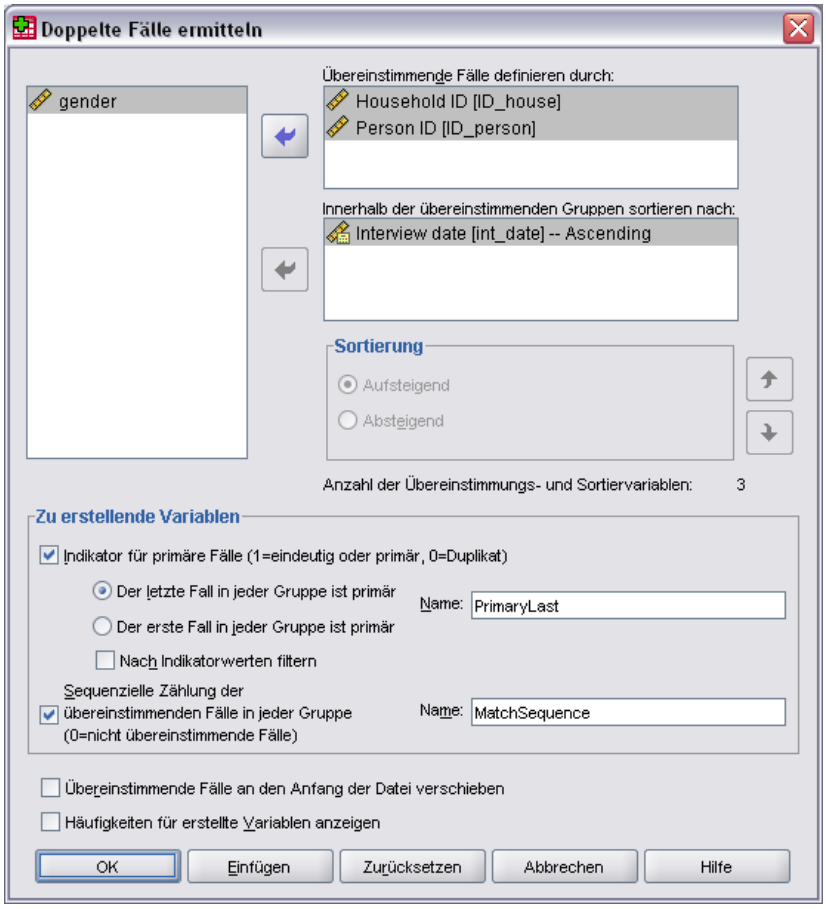
▶ Wählen Sie mindestens eine Option in der Gruppe “Zu erstellende Variablen” aus.

Die folgenden Optionen sind verfügbar:

▶ Auswahl einer oder mehrerer Variablen, um Fälle innerhalb der Gruppen zu sortieren, die durch die ausgewählten Variablen für übereinstimmende Fälle erstellt wurden. Die durch diese Variablen festgelegte Sortierreihenfolge bestimmt den “ersten” und “letzten” Fall in jeder Gruppe. Ansonsten wird die ursprüngliche Dateireihenfolge beibehalten.

▶ Automatisches Filtern doppelter Fälle, sodass sie nicht für Berichte, Diagramme oder statistische Berechnungen verwendet werden.

Abbildung 7-13
Dialogfeld "Doppelte Fälle ermitteln"



Übereinstimmende Fälle definieren durch. Fälle werden als doppelt betrachtet, wenn ihre Werte für *alle* ausgewählten Variablen übereinstimmen. Wenn Sie nur Fälle identifizieren möchten, die in jeder Hinsicht zu 100 % übereinstimmen, müssen Sie alle Variablen auswählen.

Innerhalb der übereinstimmenden Gruppen sortieren nach. Fälle werden automatisch nach den Variablen sortiert, die zur Definition übereinstimmender Fälle dienen. Sie können zusätzliche Sortiervariablen auswählen, die die Reihenfolge der Fälle innerhalb jeder Übereinstimmungsgruppe festlegen.

- Bei jeder Sortiervariablen ist eine Sortierung in aufsteigender und in absteigender Reihenfolge möglich.
- Wenn Sie mehrere Sortiervariablen auswählen, werden die Fälle nach den einzelnen Variablen in den Kategorien der vorhergehenden Variablen in der Liste sortiert. Wenn Sie zum Beispiel *Datum* als erste Sortiervariable und *Menge* als zweite auswählen, wird innerhalb der Datumskategorien nach Menge sortiert.

- Mit den nach oben und nach unten weisenden Pfeil-Schaltflächen rechts neben der Liste können Sie die Sortierreihenfolge der Variablen ändern.
- Die Sortierreihenfolge legt den "ersten" und "letzten" Fall innerhalb jeder Übereinstimmungsgruppe fest, wodurch der Wert der optionalen Indikatorvariablen für primäre Fälle bestimmt wird. Wenn Sie beispielsweise alle außer den aktuellsten Fällen in jeder Übereinstimmungsgruppe herausfiltern möchten, können Sie die Fälle innerhalb der Gruppe in aufsteigender Reihenfolge nach einer Datumsvariablen sortieren, wodurch das aktuellste Datum zum letzten Datum in der Gruppe wird.

Indikator für primäre Fälle. Erstellt eine Variable, die für alle eindeutigen Fälle und den in jeder Gruppe übereinstimmender Fälle als primären Fall identifizierten Fall den Wert 1 und für die nichtprimären doppelten Fälle in jeder Gruppe den Wert 1 annimmt.

- Der primäre Fall kann entweder der letzte oder der erste Fall in jeder Übereinstimmungsgruppe sein. Dies richtet sich nach der Sortierreihenfolge innerhalb der Gruppe. Wenn Sie keine Sortiervariablen angeben, richtet sich die Reihenfolge der Fälle innerhalb der einzelnen Gruppen nach der ursprünglichen Dateireihenfolge.
- Sie können die Indikatorvariable als **Filtervariable** verwenden, um nichtprimäre doppelte Fälle aus Berichten und Analysen auszuschließen, ohne diese Fälle aus der Datendatei zu löschen.

Sequentielle Zählung der übereinstimmenden Fälle in jeder Gruppe. Erstellt eine Variable mit einem Sequenzwert von 1 bis *n* für die Fälle innerhalb der einzelnen Übereinstimmungsgruppen. Die Sequenz beruht auf der aktuellen Reihenfolge der Fälle in jeder Gruppe. Diese ist entweder die ursprüngliche Dateireihenfolge oder die durch angegebene Sortiervariablen festgelegte Reihenfolge.

Übereinstimmende Fälle an den Anfang der Datei verschieben. Sortiert die Datendatei so, dass alle Gruppen übereinstimmender Fälle sich am Anfang der Datendatei befinden. Dadurch wird die visuelle Überprüfung der übereinstimmenden Fälle im Daten-Editor erleichtert.

Häufigkeiten für erstellte Variablen anzeigen. Häufigkeitstabellen mit Zählungen für die einzelnen Werte der erstellten Variablen. Für die Indikatorvariable für primäre Fälle beispielsweise zeigt die Tabelle die Anzahl der Fälle mit dem Wert 0 für diese Variable an, also die Anzahl der doppelten Fälle, und die Anzahl der Fälle mit dem Wert 1 für diese Variable, also die Anzahl der eindeutigen und primären Fälle.

Fehlende Werte. Bei numerischen Variablen wird der systemdefinierte fehlende Wert wie jeder andere Wert behandelt: Fälle mit dem systemdefinierten fehlenden Wert für eine ID-Variable werden so behandelt, als würden sie übereinstimmende Werte für diese Variable aufweisen. Bei String-Variablen werden Fälle ohne Wert für eine ID-Variable so behandelt, als würden sie übereinstimmende Werte für diese Variable aufweisen.

Visuelles Klassieren

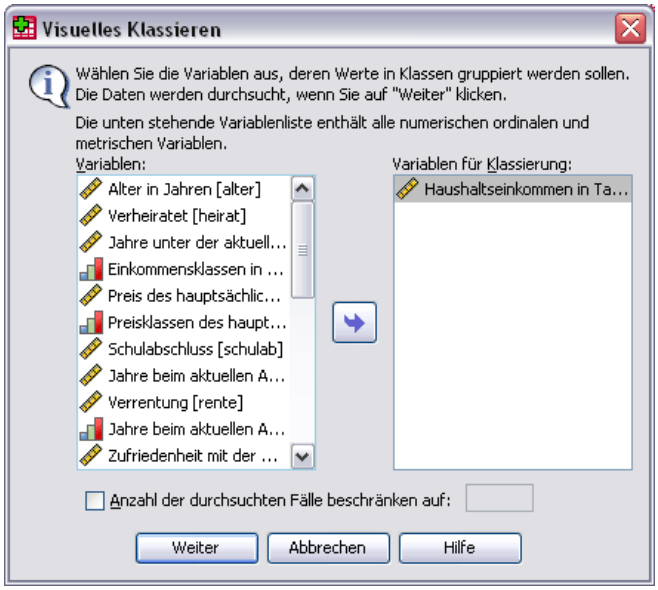
Die Funktion “Visuelles Klassieren” soll Sie beim Erstellen neuer Variablen unterstützen. Als Grundlage dafür werden nebeneinanderliegende Werte bestehender Variablen in eine begrenzte Anzahl unterschiedlicher Kategorien gruppiert. “Visuelles Klassieren” können Sie für folgende Vorgänge verwenden:

- Erstellen kategorialer Variablen aus nebeneinanderliegenden metrischen Variablen. Sie können beispielsweise eine metrische Variable für das Einkommen verwenden, um eine neue kategoriale Variable zu erstellen, die Einkommensbereiche enthält.
- Reduzieren einer großen Zahl ordinaler Kategorien zu einer kleineren Menge von Kategorien. Sie können beispielsweise eine neunstufige Bewertungsskala auf drei Kategorien, niedrig, mittel und hoch, reduzieren.

Gehen Sie im ersten Schritt wie folgt vor:

▶ Wählen Sie die numerischen metrischen und/oder ordinalen Variablen aus, für die neue kategoriale (klassierte) Variablen erstellt werden sollen.

Abbildung 7-14
Erstes Dialogfeld für die Auswahl der Variablen für das Klassieren



Wahlweise können Sie die Anzahl der zu durchsuchenden Fälle begrenzen. Bei Datendateien mit einer großen Anzahl an Fällen kann die Begrenzung der zu durchsuchenden Fälle Zeit sparen, doch Sie sollten dies nach Möglichkeit vermeiden, da es die Verteilung der Werte beeinflusst, die in den darauf folgenden Berechnungen in “Visuelles Klassieren” verwendet werden.

Anmerkung: String-Variablen und nominale numerische Variablen werden nicht in der Liste der Quellvariablen angezeigt. Für die Funktion “Visuelles Klassieren” sind numerische Variablen erforderlich, die entweder auf einem metrischen oder einem ordinalen Messniveau gemessen wurden, da die Funktion davon ausgeht, dass die Datenwerte eine logische Reihenfolge aufweisen, die für eine sinnvolle Gruppierung der Werte verwendet werden kann. In der Variablenansicht

des Daten-Editors kann das definierte Messniveau einer Variablen geändert werden. Für weitere Informationen siehe Messniveau einer Variablen in Kapitel 5 auf S. 84.

So führen Sie die Klassierung von Variablen durch:

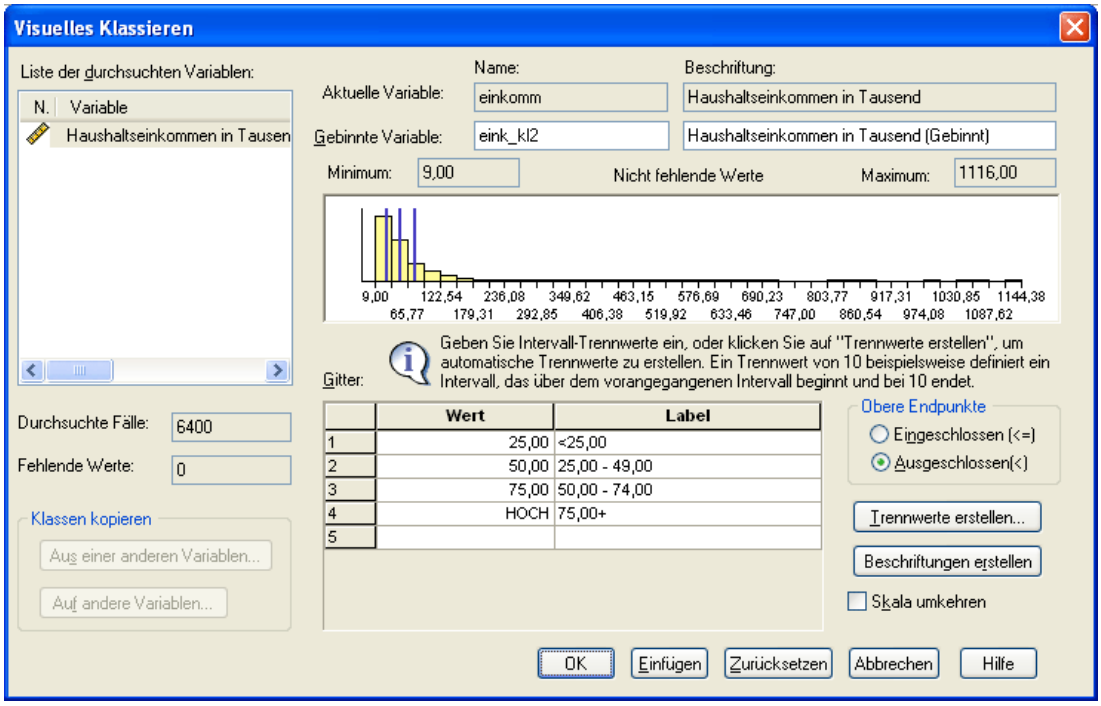
- Wählen Sie die folgenden Befehle aus den Menüs im Fenster "Daten-Editor" aus:

  Transformieren
  Visuelles Klassieren...

- Wählen Sie die numerischen metrischen und/oder ordinalen Variablen aus, für die neue kategoriale (klassierte) Variablen erstellt werden sollen.
- Wählen Sie eine Variable in der Liste der durchsuchten Variablen aus.
- Geben Sie einen Namen für die neue klassierte Variable ein. Variablennamen müssen eindeutig sein und den Regeln für Variablennamen entsprechen. Für weitere Informationen siehe Variablennamen in Kapitel 5 auf S. 83.
- Definieren Sie die Klassierungskriterien für die neue Variable. Für weitere Informationen siehe Klassieren von Variablen auf S. 132.
- Klicken Sie auf OK.

Klassieren von Variablen

Abbildung 7-15
Visuelles Klassieren, Hauptdialogfeld



Das Hauptdialogfeld von “Visuelles Klassieren” enthält folgende Informationen über die durchsuchten Variablen:

Liste der durchsuchten Variablen. Zeigt die Variablen an, die Sie im ersten Dialogfeld ausgewählt haben. Sie können die Liste anhand des Messniveaus (metrisch oder ordinal) oder anhand des Variablenlabels oder -namens sortieren, indem Sie auf die Spaltenüberschriften klicken.

Durchsuchte Fälle. Gibt die Zahl der durchsuchten Fälle an. Alle durchsuchten Fälle (ohne systemdefiniert oder benutzerdefiniert fehlende Werte) für die ausgewählte Variable werden verwendet, um die in den Berechnungen von “Visuelles Klassieren” verwendete Werteverteilung zu erstellen. Dazu gehören auch das im Hauptdialogfeld angezeigte Histogramm und Trennwerte auf der Grundlage von Perzentilen oder Einheiten der Standardabweichung.

Fehlende Werte. Gibt die Anzahl der durchsuchten Fälle mit systemdefiniert oder benutzerdefiniert fehlenden Werten an. Fehlende Werte werden in keiner der klassierten Kategorien verwendet. Für weitere Informationen siehe Benutzerdefinierte fehlende Werte in “Visuelles Klassieren” auf S. 139.

Aktuelle Variable. Der Name und das Variablenlabel (sofern vorhanden) für die derzeit ausgewählte Variable, die als Grundlage für die neue, klassierte Variable dient.

Klassierte Variable. Der Name und gegebenenfalls das Variablenlabel für die neue, klassierte Variable.

- **Name.** Sie müssen einen Namen für die neue Variable eingeben. Variablennamen müssen eindeutig sein und den Regeln für Variablennamen entsprechen. Für weitere Informationen siehe Variablennamen in Kapitel 5 auf S. 83.
- **Label.** Sie können ein aussagekräftiges Variablenlabel mit bis zu 255 Zeichen eingeben. Das Standard-Variablenlabel ist das Variablenlabel (sofern vorhanden) oder der Variablenname der Quellvariable, wobei am Ende des Labels *(Klassiert)* angehängt ist.

Minimum und Maximum. Der Mindest- und Höchstwert für die derzeit ausgewählte Variable, auf der Grundlage der durchsuchten Fälle ohne die Werte, die als benutzerdefiniert fehlend definiert wurden.

Nichtfehlende Werte. Das Histogramm zeigt die Verteilung der nichtfehlenden Werte für die derzeit ausgewählte Variable (auf der Grundlage der durchsuchten Fälle) an.

- Nach der Definition von Klassen für die neue Variable, werden im Histogramm vertikale Linien angezeigt, um die Trennwerte für die Klassendefinition anzuzeigen.
- Sie können auf die Trennwertlinien klicken und sie an andere Stellen im Histogramm ziehen, um so die Größen der Klassen zu verändern.
- Sie können Klassen entfernen, indem Sie die Trennwertlinien vom Histogramm wegziehen.

Anmerkung: Das Histogramm (mit den nichtfehlenden Werten), das Minimum und das Maximum beruhen auf den durchsuchten Werten. Wenn Sie nicht alle Fälle durchsuchen lassen, wird die tatsächliche Verteilung möglicherweise nicht richtig wiedergegeben, insbesondere, wenn die Datendatei anhand der ausgewählten Variablen sortiert wurde. Wenn Sie 0 Fälle durchsuchen, stehen keine Informationen über die Werteverteilung zur Verfügung.

Gitter. Zeigt die Werte an, die die oberen Endpunkte der einzelnen Klassen darstellen, sowie gegebenenfalls die Wertelabels für die einzelnen Klassen.

- **Wert.** Die Werte, die die oberen Endpunkte der einzelnen Klassen darstellen. Sie können Werte eingeben oder mithilfe von Trennwerte erstellen Klassen automatisch anhand ausgewählter Kriterien erstellen. Standardmäßig werden Trennwerte mit dem Wert *HOCH* automatisch aufgenommen. Diese Klasse enthält alle nichtfehlenden Werte, die über den anderen Trennwerten liegen. Die durch den untersten Trennwert definierte Klasse enthält alle nichtfehlenden Werte, die kleiner oder gleich diesem Wert sind (oder nur kleiner als dieser Wert, je nachdem, wie Sie die oberen Endpunkte definieren).
- **Label.** Optionale, aussagekräftige Labels für die Werte der neuen, klassierten Variablen. Da die Werte der neuen Variablen einfach aufeinander folgende Ganzzahlen von 1 bis *n* sind, können Labels, die angeben, wofür die Werte stehen, sehr hilfreich sein. Sie können Labels eingeben oder mithilfe von Beschriftungen erstellen automatisch Wertelabels erstellen.

So löschen Sie eine Klasse aus dem Gitter:

▶ Klicken Sie mit der rechten Maustaste entweder auf die Zelle *Wert* oder auf die Zelle *Label (Beschriftung)* für die Klasse.

▶ Wählen Sie im Kontextmenü die Option Zeile löschen.

Anmerkung: Wenn Sie die Klasse *HOCH* wählen, wird allen Fällen mit Werten, die höher sind als der letzte angegebene Trennwert, bei der neuen Variablen der Wert “Systemdefiniert fehlend” zugewiesen.

So löschen Sie alle Labels bzw. alle definierten Klassen:

▶ Klicken Sie mit der rechten Maustaste auf eine beliebige Stelle im Gitter.

▶ Wählen Sie im Kontextmenü entweder die Option Alle Beschriftungen löschen oder die Option Alle Trennwerte löschen.

Obere Endpunkte. Hiermit wird die Behandlung der Werte für die oberen Endpunkte in der Spalte *Wert* des Gitters festgelegt.

- **Eingeschlossen (<=).** Fälle mit dem in der Zelle *Wert* angegebenen Wert werden in die klassierte Kategorie aufgenommen. Wenn Sie beispielsweise die Werte 25, 50 und 75 angeben, werden Fälle mit einem Wert von exakt 25 in die erste Klasse eingeordnet, da diese alle Fälle mit Werten kleiner oder gleich 25 enthält.
- **Ausgeschlossen (<).** Fälle mit dem in der Zelle *Wert* angegebenen Wert werden nicht in die klassierte Kategorie aufgenommen. Stattdessen werden sie in die nächste Klasse aufgenommen. Wenn Sie beispielsweise die Werte 25, 50 und 75 angeben, werden Fälle mit einem Wert von exakt 25 in die zweite und nicht in die erste Klasse eingeordnet, da die erste Klasse nur Fälle mit Werten kleiner als 25 enthält.

Trennwerte erstellen. Erstellt automatisch klassierte Kategorien für Intervalle mit gleicher Breite, Intervalle mit derselben Anzahl von Fällen oder auf Standardabweichungen beruhende Intervalle. Diese Option ist nicht verfügbar, wenn 0 Fälle durchsucht wurden. Für weitere Informationen siehe Automatisches Erstellen von klassierten Kategorien auf S. 135.

Beschriftungen erstellen. Erstellt aussagekräftige Beschreibungen für die sequenziellen ganzzahligen Werte der neuen, klassierten Variablen, und zwar auf der Grundlage der Werte im Gitter und der angegebenen Behandlung der oberen Endpunkte (eingeschlossen oder ausgeschlossen).

Skala umkehren. Standardmäßig sind die Werte der neuen, klassierten Variablen aufsteigende sequenzielle Ganzzahlen von 1 bis *n*. Durch Umkehr der Skala werden die Werte zu absteigenden sequentiellen Ganzzahlen von *n* bis 1.

Klassen kopieren. Sie können die Klassierungsspezifikationen von einer anderen Variablen auf die derzeit ausgewählte Variable oder von der ausgewählten Variable auf mehrere andere Variablen kopieren. Für weitere Informationen siehe Kopieren von klassierten Kategorien auf S. 138.

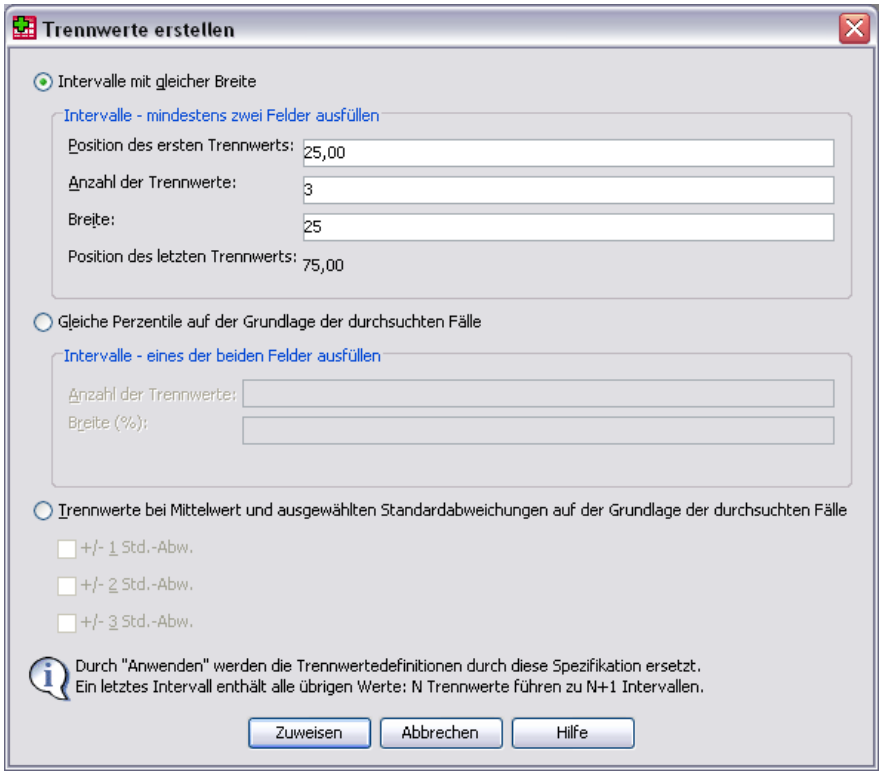
Automatisches Erstellen von klassierten Kategorien

Im Dialogfeld “Trennwerte erstellen” können Sie automatisch klassierte Kategorien erstellen, die auf den ausgewählten Kriterien beruhen.

So verwenden Sie das Dialogfeld “Trennwerte erstellen”:

- Wählen Sie (durch Klicken) eine Variable in der Liste der durchsuchten Variablen aus.
- Klicken Sie auf Trennwerte erstellen.
- Wählen Sie die Kriterien für das Erstellen der Trennwerte aus, welche die klassierten Kategorien definieren sollen.
- Klicken Sie auf Zuweisen.

Abbildung 7-16
Dialogfeld "Trennwerte erstellen"



Anmerkung: Das Dialogfeld "Trennwerte erstellen" ist nicht verfügbar, wenn 0 Fälle durchsucht wurden.

Intervalle mit gleicher Breite. Erstellt klassierte Kategorien mit gleicher Breite (z. B. 1–10, 11–20, 21–30), die auf zwei der folgenden drei Kriterien beruhen:

- **Position des ersten Trennwerts.** Der Wert, der das obere Ende der untersten klassierten Kategorie kennzeichnet (Beispiel: Der Wert 10 gibt einen Bereich an, der alle Werte bis 10 einschließt).
- **Anzahl der Trennwerte.** Die Anzahl der klassierten Kategorien ist die Anzahl der Trennwerte *plus 1*. So führen 9 Trennwerte zu 10 klassierten Kategorien.
- **Breite.** Die Breite der einzelnen Intervalle. Der Wert 10 beispielsweise teilt *Alter in Jahren* in jeweils 10 Jahre umfassende Intervalle ein.

Gleiche Perzentile auf der Grundlage der durchsuchten Fälle. Erstellt klassierte Kategorien mit der gleichen Anzahl von Fällen in jeder Klasse (unter Verwendung des **empirischen Algorithmus** für Perzentile). Als Grundlage dient eines der folgenden Kriterien:

- **Anzahl der Trennwerte.** Die Anzahl der klassierten Kategorien ist die Anzahl der Trennwerte *plus 1*. So führen drei Trennwerte zu vier Perzentilklassen (Quartilen) mit jeweils 25 % der Fälle.
- **Breite (%).** Die Breite der einzelnen Intervalle als Prozentsatz der Gesamtanzahl der Fälle. Der Wert 33,3 beispielsweise führt zu drei klassierten Kategorien (zwei Trennwerte) mit jeweils 33,3 % der Fälle.

Wenn die Quellvariable eine relativ geringe Anzahl eindeutiger Werte oder eine große Anzahl von Fällen mit demselben Wert enthält, erhalten Sie möglicherweise weniger Klassen als angefordert. Liegen mehrere identische Werte an einem Trennwert vor, werden alle Werte in dasselbe Intervall aufgenommen. Die tatsächlichen Prozentsätze sind daher nicht in jedem Fall genau gleich.

Trennwerte bei Mittelwert und ausgewählten Standardabweichungen auf der Grundlage der durchsuchten Fälle. Erstellt klassierte Kategorien auf der Grundlage der Werte für Mittelwert und Standardabweichung für die Verteilung der Variablen.

- Wenn Sie keines der Standardabweichungs-Intervalle auswählen, werden zwei klassierte Kategorien erstellt, mit dem Mittelwert als Trennwert zwischen den Klassen.
- Sie können eine beliebige Kombination von Standardabweichungs-Intervallen auf der Grundlage von einer, zwei und/oder drei Standardabweichungen auswählen. Beispiel: Wenn Sie alle drei Möglichkeiten auswählen, würde das zu 8 klassierten Kategorien führen – 6 Klassen in einem Standardabweichungs-Intervall und zwei Klassen für Fälle, die mehr als drei Standardabweichungen über bzw. unter dem Mittelwert liegen.

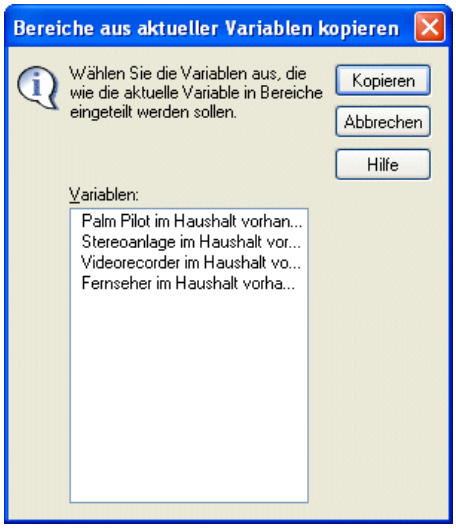
Bei einer Normalverteilung liegen 68 % der Fälle innerhalb einer Standardabweichung vom Mittelwert, 95 % innerhalb von zwei Standardabweichungen und 99 % innerhalb von drei Standardabweichungen. Das Erstellen von klassierten Kategorien auf der Grundlage von Standardabweichungen kann zu definierten Klassen außerhalb des tatsächlichen Datenbereichs und sogar außerhalb des Bereichs der möglichen Datenwerte (z. B. ein negativer Gehaltsbereich) führen.

Anmerkung: Die Berechnung von Perzentilen und Standardabweichungen beruht auf den durchsuchten Fällen. Wenn Sie die Anzahl der durchsuchten Fälle beschränken, enthalten die resultierenden Klassen möglicherweise nicht den Anteil an Fällen, den Sie in diesen Klassen wünschten, insbesondere dann, wenn die Datendatei nach der Quellvariablen sortiert wird. Beispiel: Wenn Sie nur die ersten 100 Fälle einer Datendatei mit 1000 Fällen durchsuchen lassen und die Datendatei in aufsteigender Reihenfolge nach dem Alter des Befragten sortiert ist, erhalten Sie möglicherweise nicht vier Perzentil-Altersklassen mit jeweils 25 % der Fälle, sondern in den ersten drei Klassen befinden sich vielleicht nur jeweils 3,3 % der Fälle und in der letzten Klasse 90 %.

Kopieren von klassierten Kategorien

Beim Erstellen von klassierten Kategorien für mehrere Variablen, können Sie die Klassierungsspezifikationen von einer anderen Variablen auf die derzeit ausgewählte Variable oder von der ausgewählten Variablen auf mehrere andere Variablen kopieren.

Abbildung 7-17
Kopieren von Klassen von der aktuellen Variablen bzw. auf die aktuelle Variable



So kopieren Sie Klassierungsspezifikationen:

- Definieren Sie klassierte Kategorien für mindestens eine Variable – klicken Sie jedoch *nicht* auf OK oder Einfügen.
- Wählen Sie (durch Klicken) eine Variable in der Liste der durchsuchten Variablen aus, für die Sie klassierte Kategorien definiert haben.
- Klicken Sie auf Auf andere Variablen.
- Wählen Sie die Variablen aus, für die neue Variablen mit denselben klassierten Kategorien erstellt werden sollen.
- Klicken Sie auf Kopieren.

oder

- Wählen Sie (durch Klicken) eine Variable in der Liste der durchsuchten Variablen aus, auf die Sie definierte klassierte Kategorien kopieren möchten.
- Klicken Sie auf Aus einer anderen Variablen.
- Wählen Sie die Variable mit den definierten klassierten Kategorien aus, die Sie kopieren möchten.
- Klicken Sie auf Kopieren.

Wenn Sie Wertelabels für die Variable angegeben haben, aus der Sie die Klassierungsspezifikationen kopieren, werden diese ebenfalls kopiert.

Anmerkung: Wenn Sie im Hauptdialogfeld von “Visuelles Klassieren” auf OK klicken, um die neuen klassierten Variablen zu erstellen (oder das Dialogfeld auf andere Weise schließen), können Sie “Visuelles Klassieren” nicht dazu verwenden, diese klassierten Kategorien auf andere Variablen zu verschieben.

Benutzerdefinierte fehlende Werte in “Visuelles Klassieren”

Als benutzerdefiniert fehlend definierte Werte (Werte, die als Codes für fehlende Daten gekennzeichnet wurden) für die Quellvariable werden nicht in klassierte Kategorien für die neue Variable aufgenommen. Benutzerdefinierte fehlende Werte für die Quellvariable werden als benutzerdefinierte fehlende Werte für die neue Variable kopiert, und alle definierten Wertelabels für die Codes für fehlende Werte werden ebenfalls kopiert.

Wenn ein Code für einen fehlenden Wert mit einem der Werte der klassierten Kategorien für die neue Variable in Konflikt steht, wird der Code für den fehlenden Wert für die neue Variable als nicht in Konflikt stehender Wert umkodiert, indem zum höchsten Wert einer klassierten Kategorie der Wert 100 addiert wird. Beispiel: Wenn der Wert 1 für die Quellvariable als benutzerdefiniert fehlend definiert ist und die neue Variable sechs klassierte Kategorien umfasst, haben alle Fälle mit dem Wert 1 für Quellvariable den Wert 106 für die neue Variable und 106 wird als benutzerdefiniert fehlend definiert. Wenn für den benutzerdefinierten fehlenden Wert für die Quellvariable ein Wertelabel definiert war, wird dieses Label als Wertelabel für den aufgezeichneten Wert der neuen Variablen beibehalten.

Anmerkung: Wenn die Quellvariable einen definierten Bereich benutzerdefinierter fehlender Werte der Form *LO-n* enthält (wobei *n* eine positive Zahl ist), sind die entsprechenden benutzerdefinierten fehlenden Werte für die neue Variable negative Zahlen.

Kapitel
8

Transformieren von Daten

Im Idealfall sind Ihre Rohdaten genau für die Analyse geeignet, die Sie ausführen möchten, und die Beziehungen zwischen den Variablen sind entweder linear oder rein orthogonal. Dies ist leider selten der Fall. Mit einer Vorabanalyse können problematische Kodierschemata oder Kodierfehler erkannt werden. Transformationen von Daten können auch erforderlich sein, um die tatsächliche Beziehung zwischen den Variablen herauszuarbeiten.

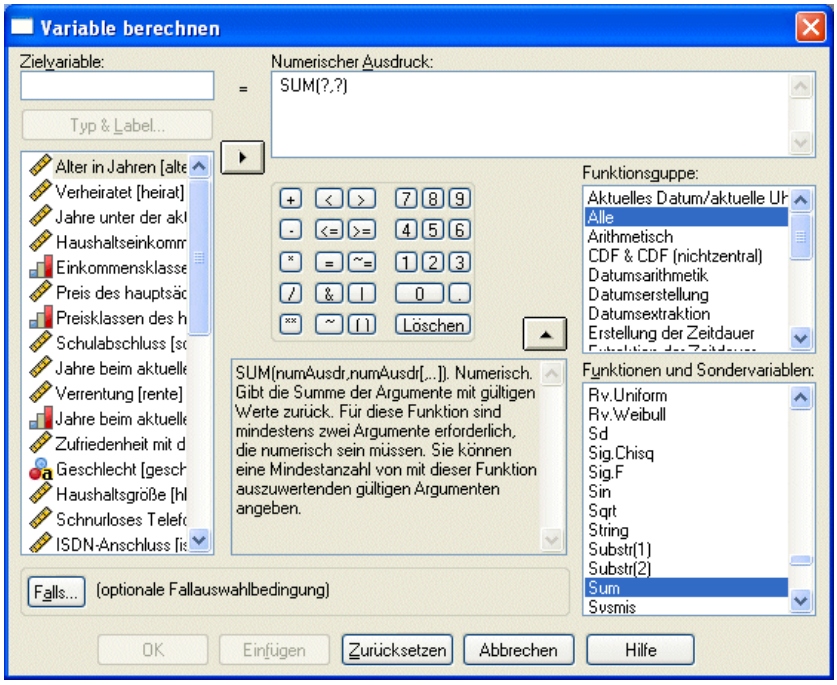
Mit SPSS können Sie verschiedene Transformationen von Daten ausführen, von einfachen Aufgaben wie dem Zusammenfassen von Kategorien zur Analyse, bis zu fortgeschritteneren Aufgaben wie dem Erstellen neuer Variablen auf der Grundlage von Bedingungen und komplizierten Gleichungen.

Berechnen von Variablen

Im Dialogfeld “Berechnen” werden Werte für Variablen auf der Grundlage von numerischen Transformationen anderer Variablen berechnet.

- Sie können Werte für numerische oder String-Variablen berechnen.
- Sie können neue Variablen erstellen oder die Werte vorhandener Variablen ersetzen. Bei neuen Variablen können Sie außerdem Variablentyp und -label angeben.
- Auf der Grundlage von logischen Bedingungen können Sie Werte für ausgewählte Teilmengen von Daten berechnen lassen.
- Sie können über 70 systemeigene Funktionen verwenden, darunter arithmetische Funktionen, Statistikfunktionen, Verteilungsfunktionen und String-Funktionen.

Abbildung 8-1
Dialogfeld "Variable berechnen"



So berechnen Sie Variablen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
Variable berechnen...

▶ Geben Sie den Namen einer einzelnen Zielvariablen ein. Dies kann eine vorhandene Variable sein oder eine neue Variable, die in die Arbeitsdatei aufgenommen werden soll.

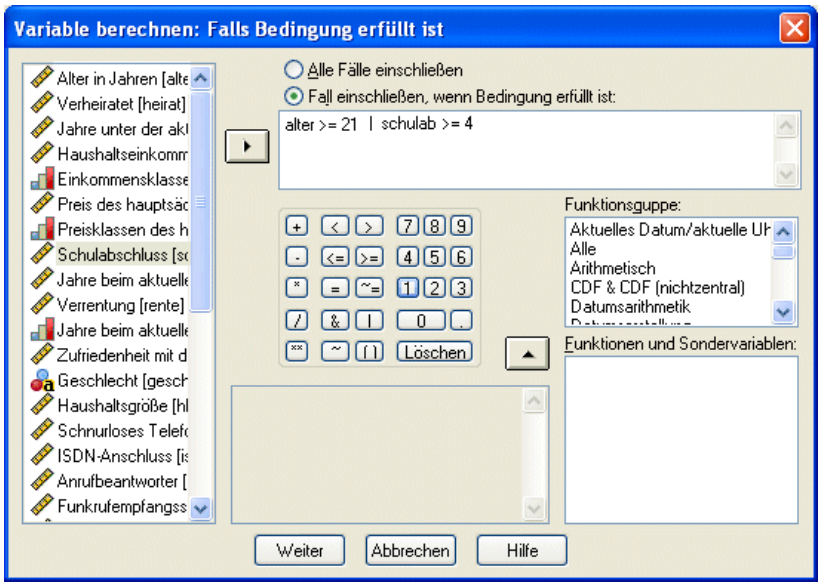
▶ Um einen Ausdruck zu bilden, fügen Sie die Elemente in das Feld "Ausdruck" ein, oder geben Sie den Ausdruck direkt in dieses Feld ein.

- Sie können Funktionen oder häufig verwendete Systemvariablen einfügen, indem Sie eine Gruppe aus der Liste "Funktionsgruppe" auswählen und in der Liste "Funktionen und Sodervariablen" auf die Funktion bzw. Variable doppelklicken (oder die Funktion bzw. Variable auswählen und auf den Pfeil neben der Liste "Funktionsgruppe" klicken). Geben Sie alle durch Fragezeichen gekennzeichneten Parameter an (gilt nur für Funktionen). Die Funktionsgruppe mit der Beschriftung Alle bietet eine Auflistung aller verfügbaren Funktionen und Systemvariablen. Eine kurze Beschreibung der aktuell ausgewählten Funktion oder Variablen wird in einem speziellen Bereich des Dialogfelds angezeigt.
- String-Konstanten müssen in Anführungszeichen oder Apostrophe eingeschlossen werden.
- Wenn die Werte Dezimalstellen enthalten, muß ein Punkt (.) als Dezimaltrennzeichen verwendet werden.
- Bei neuen String-Variablen müssen Sie außerdem Typ & Label auswählen, um den Datentyp anzugeben.

Variable berechnen: Falls Bedingung erfüllt ist

Im Dialogfeld "Variable berechnen: Falls Bedingung erfüllt ist" können Sie Transformationen auf Teilmengen von Fällen anwenden, die anhand bedingter Ausdrücke ausgewählt werden. Ein bedingter Ausdruck gibt für jeden Fall den Wert *Wahr*, *Falsch* oder *Fehlend* zurück.

Abbildung 8-2
Dialogfeld "Variable berechnen: Falls Bedingung erfüllt ist"



- Wenn das Ergebnis eines bedingten Ausdrucks *Wahr* ist, wird der Fall in die ausgewählte Untergruppe aufgenommen.
- Wenn das Ergebnis eines bedingten Ausdrucks *Falsch* oder *Fehlend* lautet, wird der Fall nicht in die ausgewählte Untergruppe aufgenommen.
- In den meisten bedingten Ausdrücke wird mindestens einer der sechs Vergleichsoperatoren (<, >, <=, >=, = und ~=) verwendet. Diese sind auf der Rechentastatur verfügbar.
- Bedingte Ausdrücke können Variablennamen, Konstanten, arithmetische Operatoren, numerische und andere Funktionen, logische Variablen und Vergleichsoperatoren enthalten.

Variable berechnen: Typ und Label

In der Standardeinstellung sind die neu berechneten Variablen numerisch. Zum Berechnen einer neuen String-Variablen müssen Sie Datentyp und Länge angeben.

Label. Optionale, aussagekräftige Variablenlabels können bis zu 255 Byte umfassen. Sie können ein Label eingeben oder die ersten 110 Zeichen des Berechnungsausdrucks als Label verwenden.

Typ. Es können numerische oder String-Variablen (alphanumerische Variablen) berechnet werden. In Berechnungen können keine String-Variablen verwendet werden.

Abbildung 8-3
Dialogfeld "Variable berechnen: Typ und Label"

Funktionen

Es werden verschiedene Typen von Funktionen unterstützt. Dazu gehören:

- Arithmetische Funktionen
- Statistische Funktionen
- String-Funktionen
- Datums- und Uhrzeitfunktionen
- Verteilungsfunktionen
- Funktionen mit Zufallsvariablen
- Funktionen mit fehlenden Werten
- Bewertungsfunktionen (nur SPSS-Server)

Weitere Informationen und eine detaillierte Beschreibung der einzelnen Funktionen erhalten Sie, wenn Sie auf der Registerkarte "Index" des Hilfesystems Funktionen eingeben.

Fehlende Werte in Funktionen

Fehlende Werte werden von Funktionen und einfachen arithmetischen Ausdrücken unterschiedlich behandelt. In dem Ausdruck:

```
(var1+var2+var3)/3
```

fehlt das Ergebnis, wenn ein Fall einen fehlenden Wert für eine der drei Variablen enthält.

In dem Ausdruck:

```
MEAN(var1,var2,var3)
```

fehlt das Ergebnis nur, wenn der Fall fehlende Werte für alle drei Variablen enthält.

Bei statistischen Funktionen können Sie die Mindestanzahl von Argumenten angeben, die nichtfehlende Werte enthalten müssen. Geben Sie dazu nach dem Namen der Funktion einen Punkt und die Mindestanzahl ein, wie zum Beispiel in

```
MEAN.2(var1,var2,var3)
```

Zufallszahlengeneratoren

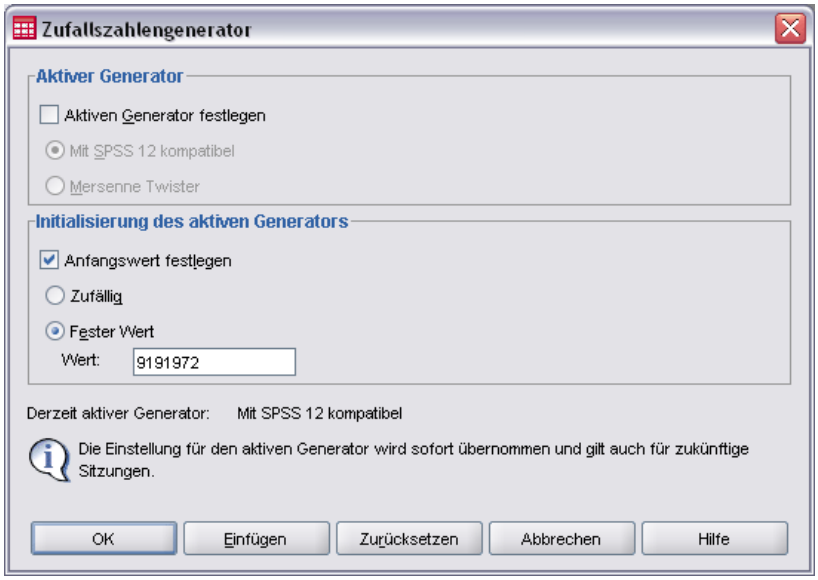
Im Dialogfeld "Zufallszahlengeneratoren" können Sie den Zufallszahlengenerator auswählen und den Startwert für eine Sequenz von Zufallszahlen festlegen.

Aktiver Generator. Zwei verschiedene Zufallszahlengeneratoren stehen zur Verfügung:

- **Version-12-kompatibel.** Der in Version 12 und früheren Versionen verwendete Zufallszahlengenerator. Wenn Sie randomisierte Ergebnisse reproduzieren möchten, die in früheren Versionen auf der Grundlage eines angegebenen Startwerts generiert wurden, müssen Sie diesen Zufallszahlengenerator verwenden.
- **Mersenne-Twister.** Ein neuerer Zufallszahlengenerator, der für Simulationszwecke eine höhere Zuverlässigkeit bietet. Sofern es nicht darum geht, zufallsbestimmte Ergebnisse aus SPSS 12 oder älteren Versionen zu reproduzieren, sollte dieser Zufallszahlengenerator verwendet werden.

Initialisierung des aktiven Generators. Der Startwert für Zufallszahlen ändert sich jedesmal, wenn SPSS eine Zufallszahl zur Verwendung bei Transformationen (wie bei den Zufallsdistributionsfunktionen), der Entnahme von Zufallsstichproben oder der Gewichtung von Fällen erzeugt. Wenn Sie eine Reihe von Zufallszahlen reproduzieren möchten, setzen Sie den Startwert auf einen bestimmten Wert zurück, bevor eine Analyse mit diesen Zufallszahlen ausgeführt wird. Dieser Wert muss eine positive ganze Zahl sein.

Abbildung 8-4
Dialogfeld "Zufallszahlengenerator"



So können Sie den Zufallszahlengenerator auswählen und/oder den Startwert festlegen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
Zufallszahlengeneratoren

Häufigkeiten von Werten in Fällen zählen

In diesem Dialogfeld wird eine Variable erstellt, mit welcher das Auftreten derselben Werte in einer Variablenliste pro Fall gezählt wird. Zum Beispiel könnte eine Befragung eine Liste von Zeitschriften mit Feldern zum Ankreuzen für *Ja* und *Nein* enthalten, mit denen die Befragten angeben, welche Zeitschriften sie lesen. Sie könnten dann die Anzahl aller Antworten mit *Ja* für jeden Befragten zählen und eine neue Variable erstellen, welche die Anzahl der gelesenen Zeitschriften enthält.

Abbildung 8-5
Dialogfeld "Häufigkeiten von Werten in Fällen zählen"

So zählen Sie die Werte in Fällen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
Werte in Fällen zählen...

▶ Geben Sie einen Namen für die Zielvariable ein.

▶ Wählen Sie mindesten zwei Variablen desselben Typs aus (numerische oder String-Variablen).

▶ Klicken Sie auf Werte definieren und geben Sie an, welcher Wert oder welche Werte gezählt werden sollen.

Wahlweise können Sie eine Teilmenge von Fällen definieren, für die Werte gezählt werden sollen.

Werte in Fällen zählen: Welche Werte?

Der Wert der Zielvariablen (im Hauptdialogfeld) wird jedesmal um 1 erhöht, wenn eine der ausgewählten Variablen einer Angabe in der Liste “Zu zählende Werte” entspricht. Wenn ein Fall auf mehrere Angaben für eine Variable zutrifft, wird die Zielvariable für diese Variablen mehrmals erhöht.

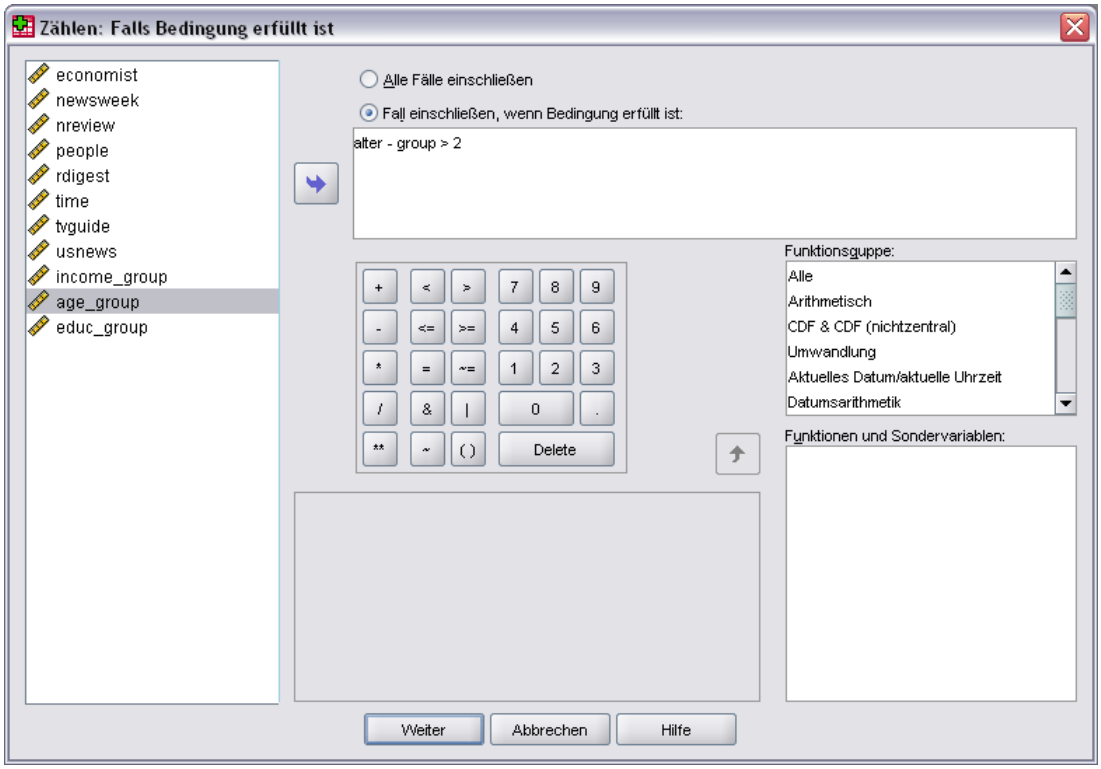
Angaben von Werten können einzelne Werte, fehlende oder systemdefinierte fehlende Werte und Bereiche enthalten. Bei Bereichen sind die Endwerte und alle benutzerdefinierten fehlenden Werte eingeschlossen, die in den Bereich fallen.

Abbildung 8-6
Dialogfeld “Werte in Fällen zählen: Welche Werte?”

Häufigkeiten von Werten in Fällen zählen: Falls Bedingung erfüllt ist

Im Dialogfeld “Variable berechnen: Falls Bedingung erfüllt ist” können Sie die Häufigkeiten von Werten für Teilmengen von Fällen zählen, die anhand bedingter Ausdrücke ausgewählt werden. Ein bedingter Ausdruck gibt für jeden Fall den Wert *Wahr*, *Falsch* oder *Fehlend* zurück.

Abbildung 8-7
Dialogfeld "Häufigkeiten von Werten in Fällen zählen: Falls Bedingung erfüllt ist"



Allgemeine Erläuterungen zur Verwendung des Dialogfelds "Falls Bedingung erfüllt ist" finden Sie unter Variable berechnen: Falls Bedingung erfüllt ist auf S. 142.

Umkodieren von Werten

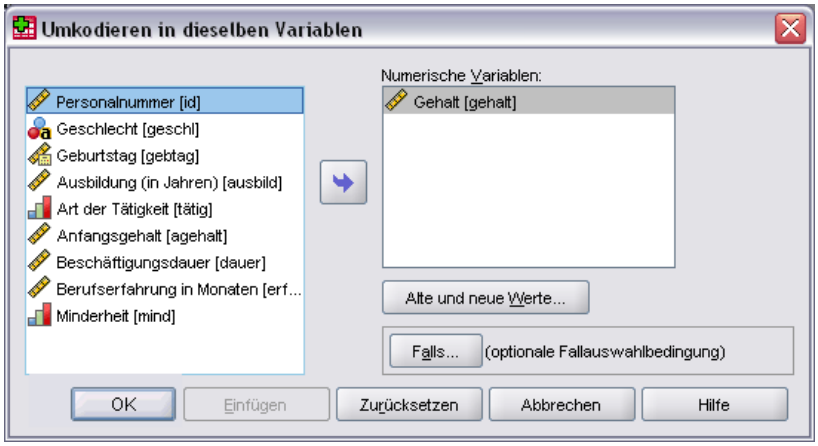
Sie können Datenwerte ändern, indem Sie diese umkodieren. Dies ist besonders nützlich, wenn Sie Kategorien zusammenfassen oder kombinieren. Sie können Werte in vorhandenen Variablen umkodieren oder neue Variablen auf der Grundlage der umkodierten Werte vorhandener Variablen erzeugen.

Umkodieren in dieselben Variablen

Im Dialogfeld "Umkodieren in dieselben Variablen" werden die Werte vorhandener Variablen erneut zugewiesen oder Bereiche vorhandener Werte in neuen Werten zusammengefaßt. So können Sie zum Beispiel Löhne in Kategorien von Lohnbereichen zusammenfassen.

Sie können numerische und String-Variablen umkodieren. Wenn Sie mehrere Variablen auswählen, müssen diese vom gleichen Typ sein. Sie können nicht numerische und String-Variablen gemeinsam umkodieren.

Abbildung 8-8
Dialogfeld "Umkodieren in dieselben Variablen"



So kodieren Sie die Werte einer Variablen um:

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Transformieren
  Umkodieren in dieselben Variablen...

- ▶ Wählen Sie die Variablen aus, die Sie umkodieren möchten. Wenn Sie mehrere Variablen auswählen, müssen diese vom gleichen Typ (numerische oder String-Variablen) sein.

- ▶ Klicken Sie auf Alte und neue Werte und geben Sie an, wie die Werte umkodiert werden sollen.

  Wahlweise können Sie eine Teilmenge von Fällen zum Umkodieren definieren. Das Dialogfeld "Falls Bedingung erfüllt ist" für diese Funktion ist mit dem entsprechenden Dialogfeld für das Zählen der Häufigkeiten identisch.

Umkodieren in dieselben Variablen: Alte und neue Werte

In diesem Dialogfeld können Sie Werte zum Umkodieren definieren. Alle angegebenen Werte müssen vom gleichen Datentyp (numerisch oder String) wie die im Hauptdialogfeld ausgewählte Variable sein.

Alter Wert. Der (die) umzukodierende(n) Wert(e). Sie können einzelne Werte, Wertebereiche und fehlende Werte umkodieren. Für String-Variablen können keine systemdefinierten fehlenden Werte und Bereiche ausgewählt werden, da keines der beiden Konzepte auf String-Variablen zutrifft. Bei Bereichen sind die Endwerte und alle benutzerdefinierten fehlenden Werte eingeschlossen, die in den Bereich fallen.

- **Wert.** Einzelner alter Wert, der in einen neuen Wert umkodiert wird. Der Wert muss vom gleichen Datentyp sein (numerisch oder String) wie die umzukodierende(n) Variable(n).
- **Systemdefiniert fehlend.** Werte, die von SPSS zugewiesen werden, wenn Werte in den Daten nicht dem festgelegten Formattyp entsprechen, numerische Felder leer sind oder ein aus einem Transformationsbefehl resultierender Wert nicht definiert ist. Numerische systemdefiniert fehlende Werte werden durch Kommata dargestellt. String-Variablen können keine systemdefiniert fehlenden Werte enthalten, da in String-Variablen alle Zeichen zulässig sind.

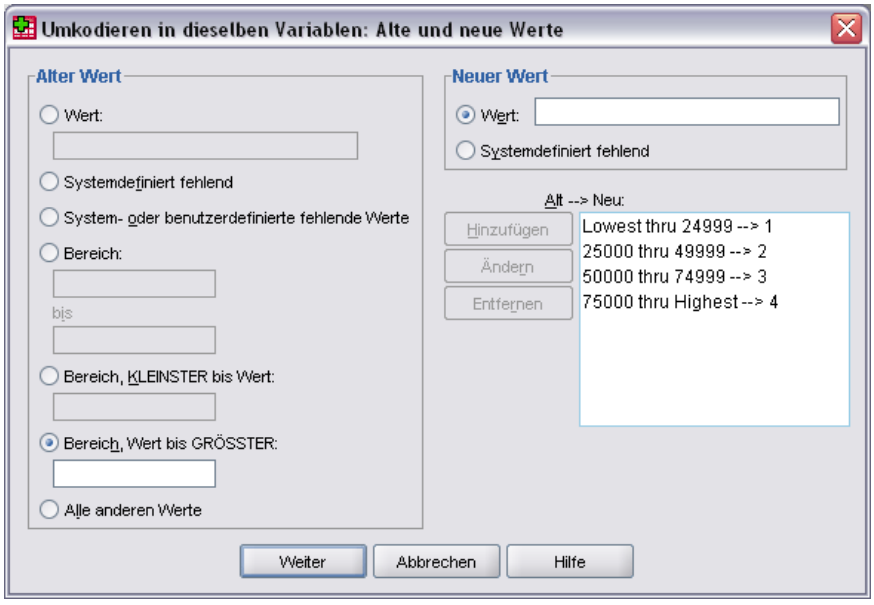
- **System- oder benutzerdefiniert fehlende Werte.** Beobachtungen mit Werten, die entweder als benutzerdefiniert fehlende Werte definiert wurden oder unbekannt sind und den systemdefiniert fehlenden Wert erhalten haben, der durch ein Komma (,) dargestellt wird.
- **Bereich.** Einschließender Wertebereich. Nicht verfügbar für String-Variablen. Alle benutzerdefinierten, fehlenden Werte innerhalb des Bereichs sind eingeschlossen.
- **Alle anderen Werte.** Alle verbleibenden Werte, die nicht in den Angaben in der Alt-Neu-Liste enthalten sind. Dies erscheint als ELSE-Anweisung in der Alt-Neu-Liste.

Neuer Wert. Der einzelne Wert, in den jeder alte Wert oder Wertebereich umkodiert wird. Sie können einen Wert eingeben oder den systemdefiniert fehlenden Wert zuweisen.

- **Wert.** Ein Wert, in den ein oder mehrere alte Werte umkodiert werden. Der Wert muss vom gleichen Datentyp (numerisch oder String) sein wie der alte Wert.
- **Systemdefiniert fehlend.** Kodiert die angegebenen alten Werte in den systemdefiniert fehlenden Wert um. Der systemdefiniert fehlende Wert wird bei Berechnungen nicht verwendet und Fälle mit dem systemdefiniert fehlenden Wert werden von vielen Prozeduren ausgeschlossen. Nicht verfügbar für String-Variablen.

Alt–>Neu. Die Liste mit den Festlegungen, die zum Umkodieren von Variablen benötigt werden. Sie können Angaben hinzufügen, ändern und aus der Liste entfernen. Die Liste wird automatisch auf der Grundlage der alten Wertangaben sortiert. Dabei wird in folgender Reihenfolge vorgegangen: einzelne Werte, fehlende Werte, Bereiche und alle anderen Werte. Wenn Sie eine Angabe zum Umkodieren in der Liste ändern, sortiert SPSS die Liste gegebenenfalls automatisch neu, um diese Reihenfolge beizubehalten.

Abbildung 8-9
Dialogfeld "Alte und neue Werte"

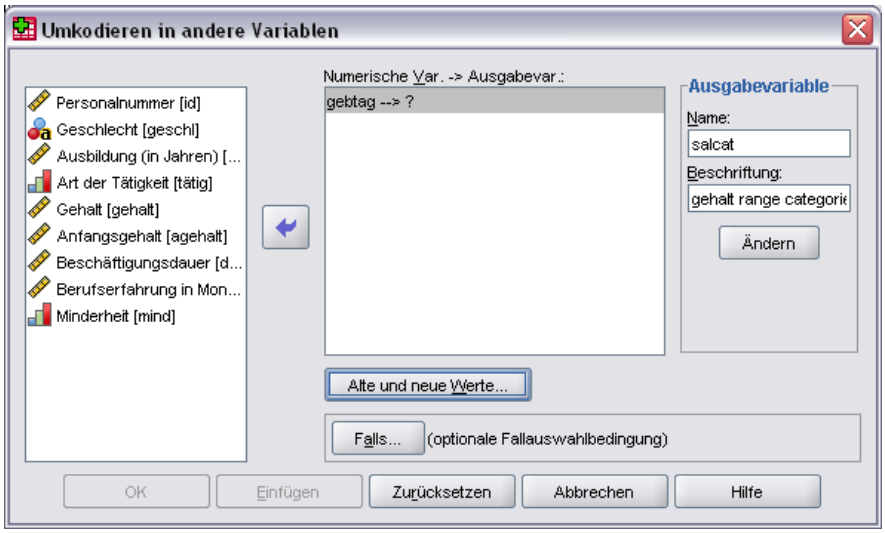


Umkodieren in andere Variablen

Im Dialogfeld “Umkodieren in andere Variablen” werden die Werte vorhandener Variablen erneut zugewiesen oder Bereiche vorhandener Werte in neuen Werten für eine neue Variable zusammengefaßt. So können Sie zum Beispiel Löhne zu einer neuen Variablen mit Lohnbereichen zusammenfassen.

- Sie können numerische und String-Variablen umkodieren.
- Sie können numerische Variablen in String-Variablen umkodieren und umgekehrt.
- Wenn Sie mehrere Variablen auswählen, müssen diese vom gleichen Typ sein. Sie können nicht numerische und String-Variablen gemeinsam umkodieren.

Abbildung 8-10
Dialogfeld “Umkodieren in andere Variablen”



So kodieren Sie die Werte einer Variablen in eine neue Variable um:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
Umkodieren in andere Variablen...

▶ Wählen Sie die Variablen aus, die Sie umkodieren möchten. Wenn Sie mehrere Variablen auswählen, müssen diese vom gleichen Typ (numerische oder String-Variablen) sein.

▶ Geben Sie für jede neue Variable einen neuen Namen an und klicken Sie auf Ändern.

▶ Klicken Sie auf Alte und neue Werte und geben Sie an, wie die Werte umkodiert werden sollen.

Wahlweise können Sie eine Teilmenge von Fällen zum Umkodieren definieren. Das Dialogfeld “Falls Bedingung erfüllt ist” für diese Funktion ist mit dem entsprechenden Dialogfeld für das Zählen der Häufigkeiten identisch.

Umkodieren in andere Variablen: Alte und neue Werte

In diesem Dialogfeld können Sie Werte zum Umkodieren definieren.

Alter Wert. Der (die) umzukodierende(n) Wert(e). Sie können einzelne Werte, Wertebereiche und fehlende Werte umkodieren. Für String-Variablen können keine systemdefinierten fehlenden Werte und Bereiche ausgewählt werden, da keines der beiden Konzepte auf String-Variablen zutrifft. Die alten Werte müssen vom gleichen Datentyp (numerisch oder String) wie die ursprüngliche Variable sein. Bei Bereichen sind die Endwerte und alle benutzerdefinierten fehlenden Werte eingeschlossen, die in den Bereich fallen.

- **Wert.** Einzelner alter Wert, der in einen neuen Wert umkodiert wird. Der Wert muss vom gleichen Datentyp sein (numerisch oder String) wie die umzukodierende(n) Variable(n).
- **Systemdefiniert fehlend.** Werte, die von SPSS zugewiesen werden, wenn Werte in den Daten nicht dem festgelegten Formattyp entsprechen, numerische Felder leer sind oder ein aus einem Transformationsbefehl resultierender Wert nicht definiert ist. Numerische systemdefiniert fehlende Werte werden durch Kommata dargestellt. String-Variablen können keine systemdefiniert fehlenden Werte enthalten, da in String-Variablen alle Zeichen zulässig sind.
- **System- oder benutzerdefiniert fehlende Werte.** Beobachtungen mit Werten, die entweder als benutzerdefiniert fehlende Werte definiert wurden oder unbekannt sind und den systemdefiniert fehlenden Wert erhalten haben, der durch ein Komma (,) dargestellt wird.
- **Bereich.** Einschließender Wertebereich. Nicht verfügbar für String-Variablen. Alle benutzerdefinierten, fehlenden Werte innerhalb des Bereichs sind eingeschlossen.
- **Alle anderen Werte.** Alle verbleibenden Werte, die nicht in den Angaben in der Alt-Neu-Liste enthalten sind. Dies erscheint als ELSE-Anweisung in der Alt-Neu-Liste.

Neuer Wert. Der einzelne Wert, in den jeder alte Wert oder Wertebereich umkodiert wird. Die neuen Werte können numerische oder String-Variablen sein.

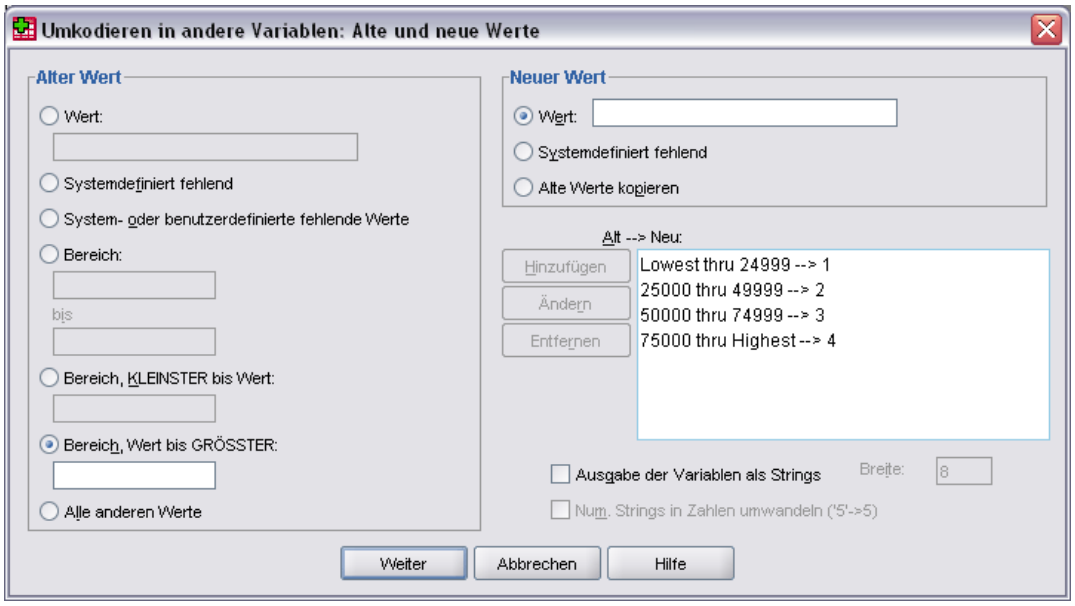
- **Wert.** Ein Wert, in den ein oder mehrere alte Werte umkodiert werden. Der Wert muss vom gleichen Datentyp (numerisch oder String) sein wie der alte Wert.
- **Systemdefiniert fehlend.** Kodiert die angegebenen alten Werte in den systemdefiniert fehlenden Wert um. Der systemdefiniert fehlende Wert wird bei Berechnungen nicht verwendet und Fälle mit dem systemdefiniert fehlenden Wert werden von vielen Prozeduren ausgeschlossen. Nicht verfügbar für String-Variablen.
- **Kopieren alter Werte.** Behält den alten Wert bei. Falls einige Werte keine Umkodierung benötigen, können Sie mit dieser Funktion die alten Werte einschließen. Alle alten Werte, die nicht angegeben wurden, werden nicht in den neuen Variablen eingeschlossen und Fällen mit diesen Werten wird der systemdefiniert fehlende Wert für die neue Variable zugewiesen.

Ausgabe der Variablen als Strings. Definiert die neue, umkodierte Variable als String-Variable (alphanumerische Variable). Die alte Variable kann eine numerische Variable oder eine String-Variable sein.

Umwandeln numerischer Strings in Zahlen. Konvertiert String-Werte, die Zahlen enthalten, in numerische Werte. Strings, die andere Zeichen als Zahlen und ein optionales Vorzeichen (+ oder -) enthalten, wird der systemdefiniert fehlende Wert zugewiesen.

Alt–>Neu. Die Liste mit den Festlegungen, die zum Umkodieren von Variablen benötigt werden. Sie können Angaben hinzufügen, ändern und aus der Liste entfernen. Die Liste wird automatisch auf der Grundlage der alten Wertangaben sortiert. Dabei wird in folgender Reihenfolge vorgegangen: einzelne Werte, fehlende Werte, Bereiche und alle anderen Werte. Wenn Sie eine Angabe zum Umkodieren in der Liste ändern, sortiert SPSS die Liste gegebenenfalls automatisch neu, um diese Reihenfolge beizubehalten.

Abbildung 8-11
Dialogfeld "Alte und neue Werte"



Rangfolge bilden

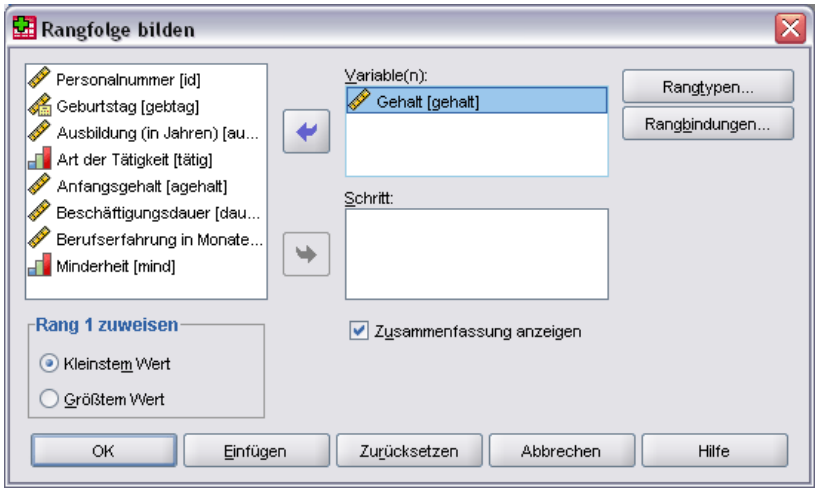
Im Dialogfeld "Fällen Rang zuordnen" werden für numerische Variablen neue Variablen mit Rängen, Normalrangwerten, Savage-Werten und Perzentil-Werten erstellt.

Neue Variablennamen und beschreibende Variablenlabels werden von SPSS automatisch auf der Grundlage des ursprünglichen Variablennamens und der ausgewählten Maße erzeugt. In einer Übersichtstabelle werden die ursprünglichen Variablen, die neuen Variablen und die Variablenlabels aufgelistet.

Die folgenden Optionen sind verfügbar:

- Rangfolge der Fälle in aufsteigender oder absteigender Reihenfolge bilden.
- Rangfolgen in Untergruppen organisieren, indem Sie mindestens eine Gruppenvariable für die Liste "Nach" auswählen. Ränge werden innerhalb jeder Gruppe berechnet. Gruppen werden durch die Kombination der Werte der Gruppenvariablen definiert. Wenn Sie zum Beispiel *Geschlecht* und *Minderheit* als Gruppenvariablen auswählen, werden Ränge für jede Kombination von *Geschlecht* und *Minderheit* berechnet.

Abbildung 8-12
Dialogfeld "Rangfolge bilden"



So bilden Sie die Rangfolge der Fälle:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Transformieren
  Rangfolge bilden...

- Wählen Sie mindestens eine Variable, für die Sie eine Rangfolge bilden möchten. Sie können nur Rangfolgen von numerischen Variablen bilden.

  Sie können die Ränge der Fälle wahlweise in aufsteigender oder absteigender Reihenfolge anordnen und Ränge in Untergruppen organisieren.

Rangfolge bilden: Typen

Zum Bilden der Rangfolge stehen mehrere Methoden zur Verfügung. Für jede Methode wird eine separate Rangvariable erstellt. Beim Bilden der Rangfolge können einfache Ränge, Savage-Werte, relative Ränge und Perzentile eingesetzt werden. Sie können die Rangfolge auch auf der Grundlage von Anteilsschätzungen und Normalrangwerten erzeugen.

Rang. Einfacher Rang. Der Wert der neuen Variablen ist gleich ihrem Rang.

Savage-Wert. Die neue Variable enthält Savage-Werte auf der Grundlage einer exponentiellen Verteilung.

Relative Rangfolge. Der Wert der neuen Variablen ist gleich dem Rangwert geteilt durch die Summe der Gewichtungen nichtfehlender Fälle.

Relative Rangfolge in Prozent. Jeder Rangwert wird geteilt durch die Anzahl der Fälle mit gültigen Werten und multipliziert mit 100.

Summe der Fallgewichtungen. Der Wert der neuen Variablen ist gleich der Summe der Fallgewichte. Die neue Variable ist für alle Fälle in derselben Gruppe eine Konstante.

N-Perzentile. Ränge basieren auf Perzentilgruppen, wobei jede Gruppe ungefähr die gleiche Anzahl von Fällen enthält. So erhalten beispielsweise bei 4 N-Perzentilen Fälle unter dem 25. Perzentil den Rang 1, Fälle zwischen dem 25. und 50. Perzentil den Rang 2, Fälle zwischen dem 50. und 75. Perzentil den Rang 3 und Fälle über dem 75. Perzentil den Rang 4.

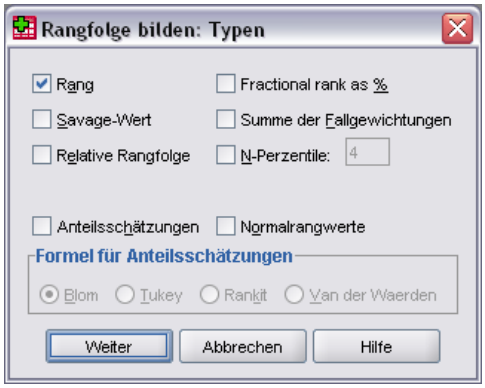
Anteilsschätzungen. Anteilsschätzer sind Schätzungen des kumulierten Anteils der Verteilung bezüglich eines einzelnen Ranges.

Normalrangwerte. Die Z-Werte, welche dem geschätzten kumulativen Anteil entsprechen.

Formel für Anteilsschätzungen. Für Anteilsschätzungen und Normalrangwerte können Sie die Formel für die Anteilsschätzung auswählen: Blom, Tukey, Rankit oder Van der Waerden.

- **Blom.** Erstellt eine neue Rangvariable auf der Grundlage von mit der Formel (r-3/8) / (w+1/4) berechneten Anteilsschätzern, wobei w die Summe der Fallgewichtungen und r der Rang ist.
- **Tukey.** Verwendet die Formel (r-1/3) / (w+1/3), wobei r den Rang und w die Summe der Fallgewichte angibt.
- **Rankit.** Es wird die Formel (r-1/2) / w verwendet, wobei w die Anzahl der Beobachtungen und r der Rang ist, der von 1 bis w reicht.
- **Van der Waerden.** Durch die Formel r/(w+1) definierte Van-der-Waerden-Transformation, wobei w die Summe der Fallgewichte und r den von 1 bis n reichenden Rang darstellt.

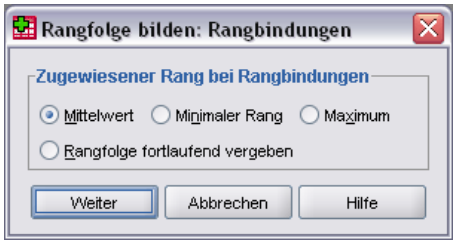
Abbildung 8-13
Dialogfeld "Rangfolge bilden: Typen"



Rangfolge bilden: Bindungen

In diesem Dialogfeld werden Einstellungen für die Methode zum Zuweisen von Rängen zu Fällen mit demselben Wert in der ursprünglichen Variablen vorgenommen.

Abbildung 8-14
Dialogfeld "Rangfolge bilden: Rangbindungen"



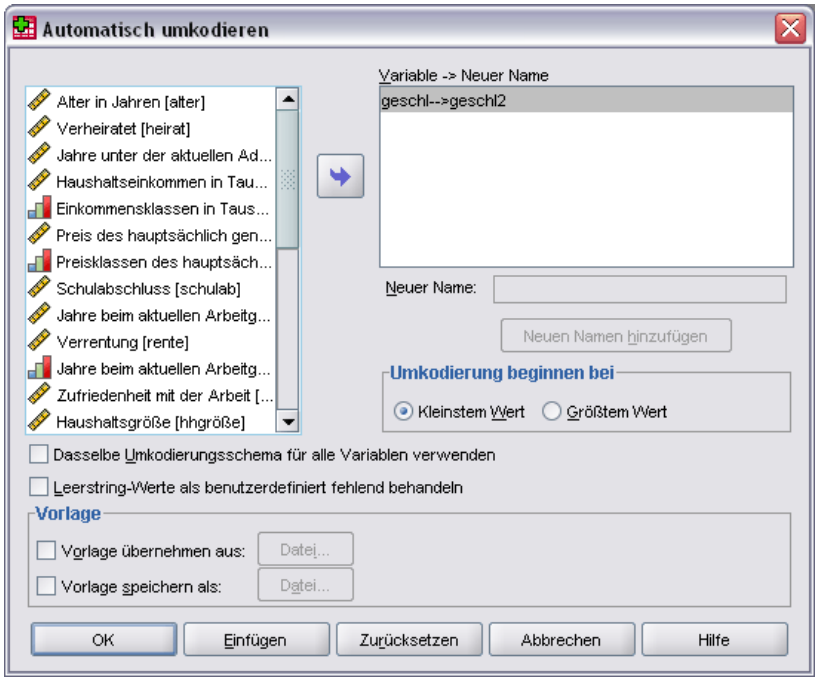
Die folgende Tabelle zeigt, wie den gebundenen Werten bei verschiedenen Methoden Ränge zugewiesen werden.

Wert	Mittelwert	Kleinster Wert	Größter Wert	Fortlaufend
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Automatisch umkodieren

Im Dialogfeld "Automatisch umkodieren" wandeln Sie String-Werte und numerische Werte in fortlaufende Ganzzahlen um. Wenn Kategoriecodes nicht sequentiell sind, vermindern die daraus resultierenden leeren Zellen die Leistung und erhöhen den Speicherbedarf für viele SPSS-Prozeduren. Außerdem können einige Prozeduren keine String-Variablen verwenden, und einige erfordern aufeinander folgende ganzzahlige Werte als Faktorstufen.

Abbildung 8-15
Dialogfeld "Automatisch umkodieren"



- Bei den mit "Automatisch umkodieren" erstellten neuen Variablen werden alle definierten Variablen und Wertelabels aus der alten Variablen beibehalten. Bei allen Werten ohne definiertes Wertelabel wird der ursprüngliche Wert als Label für den umkodierten Wert verwendet. Die alten und neuen Werte und Wertelabels werden in einer Tabelle angezeigt.
- String-Werte werden in alphabetischer Reihenfolge umkodiert, wobei Großbuchstaben jeweils vor den entsprechenden Kleinbuchstaben stehen.
- Fehlende Werte werden unter Beibehaltung ihrer Reihenfolge in fehlende Werte umkodiert, die größer als alle nichtfehlenden Werte sind. Wenn zum Beispiel die ursprüngliche Variable über 10 nichtfehlende Werte verfügt, würde der kleinste fehlende Wert auf 11 umkodiert, und der Wert 11 wäre ein fehlender Wert für die neue Variable.

Dasselbe Umkodierungsschema für alle Variablen verwenden. Mit dieser Option können Sie ein einziges Schema für die automatische Umkodierung auf alle ausgewählten Variablen anwenden und so ein einheitliches Kodierungsschema für alle neuen Variablen erzielen.

Bei dieser Option gelten die folgenden Regeln und Einschränkungen:

- Alle Variablen müssen denselben Typ aufweisen (numerische Variable oder String-Variable).
- Alle beobachteten Werte für alle ausgewählten Variablen dienen als Grundlage für eine Sortierreihenfolge, mit der die Werte vor der Umkodierung in fortlaufende Ganzzahlen sortiert werden.
- Benutzerdefiniert fehlende Werte für die neuen Variablen beruhen auf der ersten Variable in der Liste mit angegebenen benutzerdefiniert fehlenden Werten. Alle anderen Werte aus anderen ursprünglichen Variablen (mit Ausnahme der systemdefiniert fehlenden Werte) werden als gültige Werte behandelt.

Leerstring-Werte als benutzerdefiniert fehlend behandeln. Bei String-Variablen werden leere Werte oder Null-Werte nicht als **systemdefiniert fehlend** behandelt. Mit dieser Option werden leere Strings automatisch in einen **benutzerdefiniert fehlenden** Wert umkodiert, der höher ist als der höchste nichtfehlende Wert.

Vorlagen zum Definieren von Variablen

Sie können das Schema für die automatische Kodierung in einer Vorlagendatei speichern und dann auf andere Variablen und andere Datendateien anwenden.

Sie verwenden beispielsweise zahlreiche alphanumerische Produktcodes, die Sie jeden Monat automatisch in Ganzzahlen umkodieren lassen. In einigen Monaten werden jedoch neue Produktcodes eingeführt, die das ursprüngliche Schema für die automatische Umkodierung ändern. Wenn Sie das ursprüngliche Schema in einer Vorlage speichern und dann auf die neuen Daten anwenden, die die neuen Codes enthalten, werden alle neuen Codes in den Daten automatisch in Werte umkodiert, die höher sind als der höchste Wert in der Vorlage. Auf diese Weise wird das ursprüngliche Schema für die automatische Umkodierung der ursprünglichen Produktcodes beibehalten.

Vorlage speichern als. Speichert das Schema für die automatische Umkodierung der ausgewählten Variablen in einer externen Vorlagendatei.

- Mit den Informationen in der Vorlage werden die ursprünglichen nichtfehlenden Werte den umkodierten Werten zugeordnet.
- Nur Informationen für nichtfehlende Werte werden in der Vorlage gespeichert. Informationen zu benutzerdefiniert fehlenden Werten werden nicht beibehalten.
- Wenn Sie mehrere Variablen für die Umkodierung ausgewählt haben, ohne dabei dasselbe Schema für die automatische Umkodierung für alle Variablen festzulegen oder wenn keine vorhandene Vorlage im Rahmen der automatischen anzuwenden, wird die Vorlage auf der Grundlage der ersten Variable in der Liste aufgebaut.
- Wenn Sie mehrere Variablen für die Umkodierung ausgewählt und dabei die Option Dasselbe Umkodierungsschema für alle Variablen verwenden und/oder die Option Vorlage zuweisen aktiviert haben, enthält die Vorlage das kombinierte Schema für die automatische Umkodierung für alle Variablen.

Vorlage übernehmen aus. Wendet eine zuvor gespeicherte Vorlage für die automatische Umkodierung auf alle Variablen an, die zur Umkodierung ausgewählt wurden. Alle zusätzlich in den Variablen gefundenen Werte werden an das Ende des Schemas angehängt und die Beziehung zwischen den ursprünglichen und den automatisch umkodierten Werten im gespeicherten Schema bleibt erhalten.

- Alle zur Umkodierung ausgewählten Variablen müssen denselben Typ aufweisen (numerische Variable oder String-Variable), und dieser Typ muss mit dem Typ übereinstimmen, der in der Vorlage definiert ist.
- Vorlagen enthalten keine Informationen zu benutzerdefiniert fehlenden Werten. Benutzerdefiniert fehlende Werte für die Zielvariablen beruhen auf der ersten Variable in der Liste mit angegebenen benutzerdefiniert fehlenden Werten. Alle anderen Werte aus anderen ursprünglichen Variablen (mit Ausnahme der systemdefiniert fehlenden Werte) werden als gültige Werte behandelt.

- Wertzuordnungen aus der Vorlage werden als erstes angewendet. Alle verbleibenden Werte werden in Werte umkodiert, die höher sind als der letzte Wert in der Vorlage. Benutzerdefiniert fehlende Werte (auf der Grundlage der ersten Variable in der Liste mit definierten benutzerdefiniert fehlenden Werten) werden dabei in Werte umkodiert, die höher sind als der letzte gültige Wert.
- Wenn Sie mehrere Variablen für die automatische Umkodierung ausgewählt haben, wird zunächst die Vorlage angewendet. Anschließend wird eine kombinierte automatische Standard-Umkodierung für alle zusätzlichen Werte für die ausgewählten Variablen ausgeführt. So entsteht ein einziges gemeinsames Schema für die automatische Umkodierung, das alle ausgewählten Variablen erfaßt.

So kodieren Sie String- oder numerische Werte in fortlaufende Ganzzahlen um:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
  Automatisch umkodieren

▶ Wählen Sie mindestens eine Variable zum Umkodieren aus.

▶ Geben Sie für jede ausgewählte Variable einen Namen für die neue Variable ein und klicken Sie auf Neuer Name.

Assistent für Datum und Uhrzeit

Der Assistent für Datum und Uhrzeit vereinfacht eine Reihe von Aufgaben im Zusammenhang mit Datums- und Zeitvariablen.

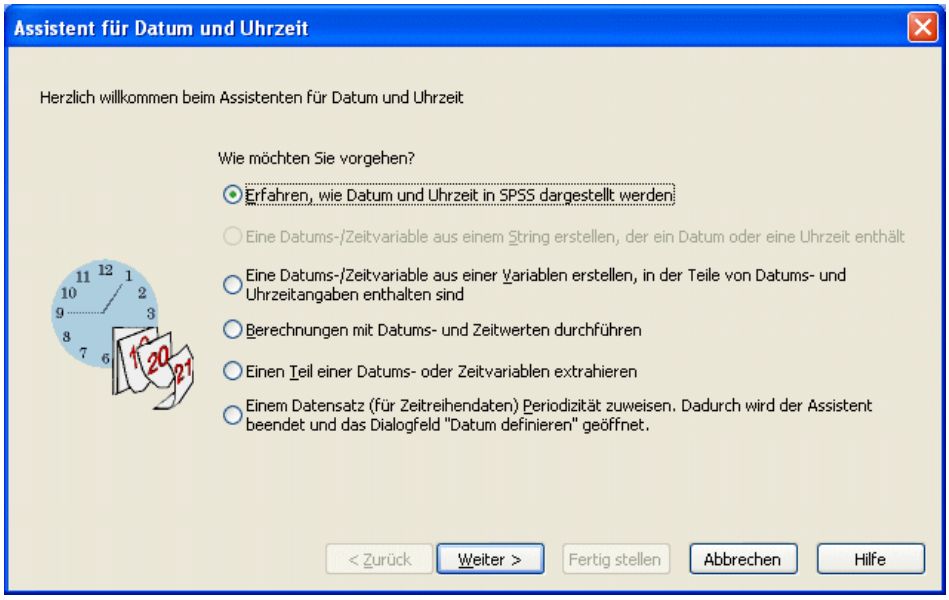
So verwenden Sie den Assistenten für Datum und Uhrzeit:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
  Assistent für Datum und Uhrzeit...

▶ Wählen Sie die gewünschte Aufgabe aus, und befolgen Sie die Schritte zur Definition der Aufgabe.

Abbildung 8-16
Assistent für Datum und Uhrzeit: Einführungsbildschirm



- **Erfahren, wie Datum und Uhrzeit dargestellt werden.** Diese Option führt Sie zu einem Bildschirm, in dem ein kurzer Überblick über die Datums-/Zeitvariablen in SPSS geboten wird. Wenn Sie auf Hilfe klicken, wird außerdem eine Verknüpfung zu detaillierteren Informationen angezeigt.
- **Eine Datums-/Zeitvariable aus einem String erstellen, der ein Datum oder eine Uhrzeit enthält.** Mit dieser Option können Sie eine Datums-/Zeitvariable aus einer String-Variablen erstellen. Beispiel: Sie haben eine String-Variable mit Datumsangaben im Format mm/tt/jjjj und möchten daraus eine Datums-/Zeitvariable erstellen.
- **Eine Datums-/Zeitvariable aus einer Variablen erstellen, in der Teile von Datums- und Uhrzeitangaben enthalten sind.** Mit dieser Option können Sie eine Datums-/Zeitvariable aus einem Set bestehender Variablen erstellen. Beispiel: Sie haben eine Variable für den Monat (als ganze Zahl), eine weitere Variable für den Tag im Monat und eine dritte für das Jahr. Sie können diese drei Variablen zu einer einzigen Datums-/Zeitvariablen kombinieren.
- **Berechnungen mit Datums- und Zeitwerten durchführen.** Diese Option dient zum Addieren oder Subtrahieren von Werten zu bzw. von Datums-/Zeitvariablen. Beispiel: Sie können die Dauer eines Prozesses berechnen, indem Sie eine Variable, die die Startzeit des Prozesses angibt, von einer anderen Variablen subtrahieren, die den Zeitpunkt des Prozeßendes angibt.
- **Einen Teil einer Datums- oder Zeitvariablen extrahieren.** Mit dieser Option können Sie einen Teil einer Datums-/Zeitvariablen extrahieren, beispielsweise des Tages im Monat aus einer Datums-/Zeitvariablen mit dem Format mm/tt/jjjj.
- **Einem Datensatz Periodizität zuweisen.** Mit dieser Option gelangen Sie zum Dialogfeld “Datum definieren”, das zum Erstellen von Datums-/Zeitvariablen verwendet wird, die aus einer Reihe aufeinander folgender Datumsangaben bestehen. Diese Funktion wird normalerweise verwendet, um Datumsangaben Zeitreihendaten zuzuweisen.

Anmerkung: Aufgaben werden deaktiviert, wenn das Daten-Set nicht die für die Ausführung der Aufgabe erforderlichen Variablen aufweist. Wenn das Daten-Set beispielsweise keine String-Variablen enthält, findet die Aufgabe zur Erstellung einer Datums-/Zeitvariablen aus einer String-Variablen keine Anwendung und ist deaktiviert.

Datums- und Zeitangaben in SPSS

Variablen für Datums- und Zeitangaben in SPSS weisen einen numerischen Variablentyp auf, mit Anzeigeformaten die den jeweiligen Datums-/Zeitformaten entsprechen. Diese Variablen werden im allgemeinen als Datums-/Zeitvariablen bezeichnet. Es wird zwischen Datums-/Zeitvariablen unterschieden, die tatsächlich für einen bestimmten Datumswert stehen, und solchen, die eine Zeitdauer repräsentieren, die unabhängig von einem bestimmten Datum ist, wie beispielsweise 20 Stunden, 10 Minuten und 15 Sekunden. Letztere werden als Dauer-Variablen und erster als Datums- oder Datums-/Zeitvariablen bezeichnet. Eine vollständige Liste der Anzeigeformate finden Sie in der *Command Syntax Reference*, im Abschnitt "Universals" unter "Date and Time".

Datums- und Datums-/Zeitvariablen. Datumsvariablen weisen ein Format auf, das einem Datum entspricht, beispielsweise mm/tt/jjjj. Datums-/Zeitvariablen weisen ein Format auf, das einem Datum und einer Uhrzeit entspricht, beispielsweise tt-mmm-jjjj hh:mm:ss. Intern werden Datums- und Datums-/Zeitvariablen als die Anzahl der seit dem 14. Oktober 1582 vergangenen Sekunden gespeichert. Datums- und Datums-/Zeitvariablen werden manchmal als Variablen mit Datumsformat bezeichnet.

- Jahresangaben werden in zweistelligen und im vierstelligen Format erkannt. In der Standardeinstellung wird bei zweistelligen Jahreszahlen ein Bereich angenommen, der 69 Jahre vor dem gegenwärtigen Datum und 30 Jahre danach umfasst. Dieser Bereich hängt von den Optionseinstellungen ab und kann konfiguriert werden (wählen Sie dazu im Menü "Bearbeiten" den Befehl Optionen aus und klicken Sie auf die Registerkarte Daten).
- In Formaten vom Typ Tag-Monat-Jahr können Bindestriche, Punkte, Kommata, Schrägstriche und Leerzeichen als Trennzeichen verwendet werden.
- Monate können durch arabische oder römische Ziffern und aus drei Buchstaben bestehende Abkürzungen dargestellt oder vollständig ausgeschrieben werden. Abkürzungen aus drei Buchstaben und vollständig ausgeschriebene Monatsnamen müssen in englischer Sprache vorliegen; Monatsnamen in anderen Sprachen werden nicht erkannt.

Dauer-Variablen. Dauer-Variablen weisen ein Format auf, das einer Zeitdauer entspricht, beispielsweise hh:mm. Sie werden intern als Sekunden ohne Bezug auf ein bestimmtes Datum gespeichert.

- Bei Zeitangaben (gilt für Datums-/Zeit- und Dauer-Variablen) können Doppelpunkte als Trennzeichen zwischen Stunden, Minuten und Sekunden verwendet werden. Stunden und Minuten sind erforderlich, Sekunden dagegen sind optional. Ein Punkt ist erforderlich, um Sekunden von Sekundenbruchteilen zu trennen. Für Stunden kann ein beliebig hoher Wert angegeben werden. Der maximale Wert für die Minuten ist jedoch 59 und für die Sekunden 59,999...

Aktuelles Datum und aktuelle Uhrzeit. Die Systemvariable *$TIME* enthält das aktuelle Datum und die aktuelle Uhrzeit. Sie steht für die Anzahl der Sekunden seit dem 14. Oktober 1582 bis zu dem Datum und der Uhrzeit, zu der der Transformationsbefehl, der diese Variable verwendet, ausgeführt wird.

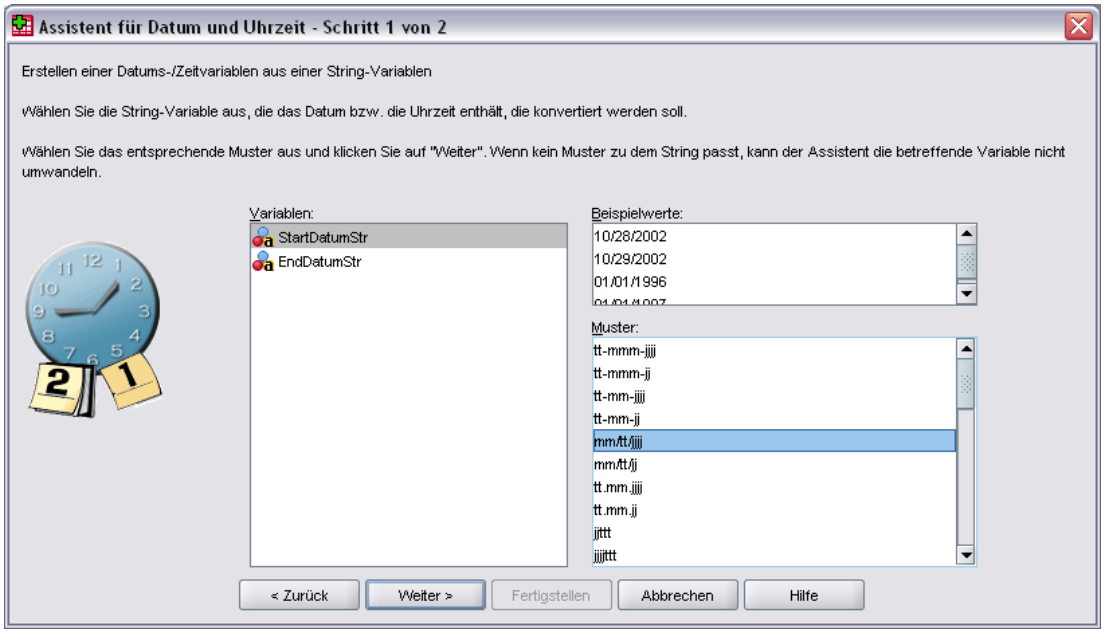
Erstellen einer Datums-/Zeitvariablen aus einer String-Variablen

So erstellen Sie eine Datums-/Zeitvariable aus einer String-Variablen:

- Wählen Sie im Einführungsbildschirm des Assistenten für Datum und Uhrzeit die Option Eine Datums-/Zeitvariable aus einem String erstellen, der ein Datum oder eine Uhrzeit enthält aus.

Auswählen einer String-Variablen zur Umwandlung in eine Datums-/Zeitvariable

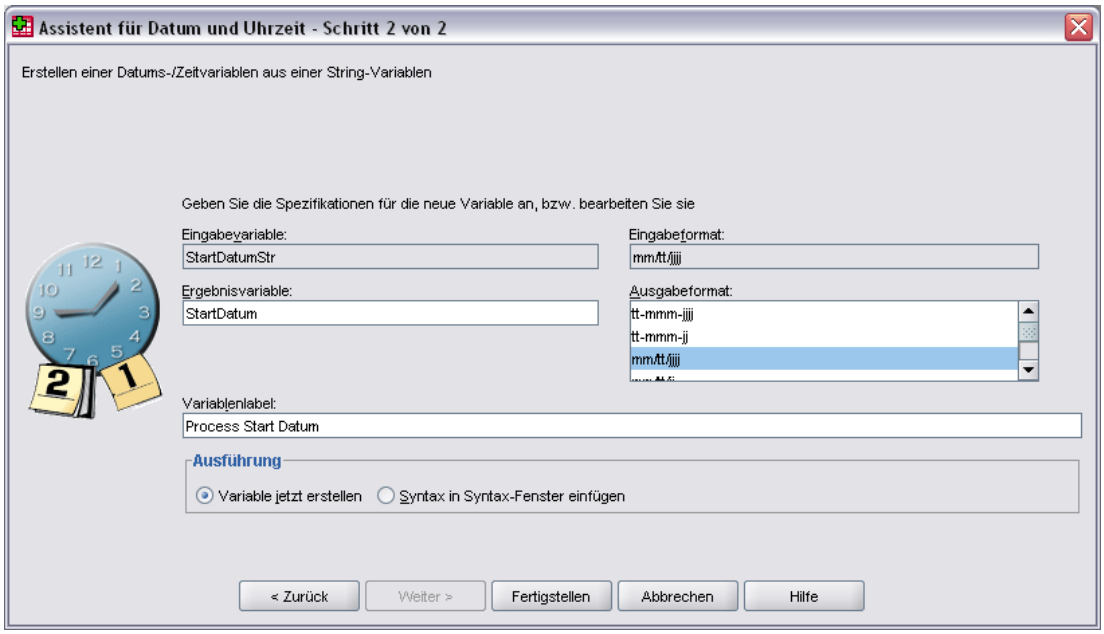
Abbildung 8-17
Erstellen einer Datums-/Zeitvariablen aus einer String-Variablen, Schritt 1



- Wählen Sie die umzuwandelnde String-Variable in der Variablenliste aus. Beachten Sie, dass in der Liste nur String-Variablen angezeigt werden.
- Wählen Sie aus der Liste “Muster” das Muster aus, das mit der Form der Datumsangaben in der String-Variablen übereinstimmt. In der Liste “Beispielwerte” werden die tatsächlichen Werte der ausgewählten Variablen in der Datendatei angezeigt. Werte der Stringvariablen, die nicht zum ausgewählten Muster passen, führen zum Wert “systemdefiniert fehlend” für die neue Variable.

Angeben des Ergebnisses der Umwandlung einer String-Variablen in eine Datums-/Zeitvariable

Abbildung 8-18
Erstellen einer Datums-/Zeitvariablen aus einer String-Variablen, Schritt 2



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.

  Die folgenden Optionen sind verfügbar:

  - Auswählen eines Datums-/Uhrzeitformats für die neue Variable in der Liste “Ausgabeformat”
  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

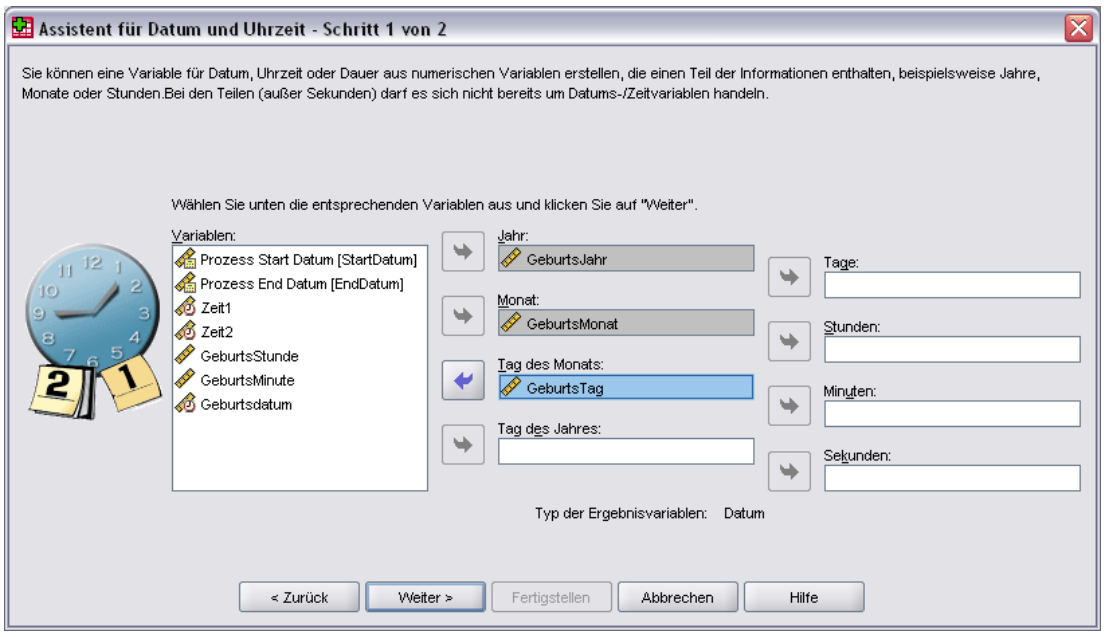
Erstellen einer Datums-/Zeitvariablen aus einem Variablen-Set

So führen Sie eine Menge bestehender Variablen zu einer einzigen Datums-/Zeitvariablen zusammen:

- Wählen Sie im Einführungsbildschirm des Assistenten für Datum und Uhrzeit die Option Eine Datums-/Zeitvariable aus einer Variablen erstellen, in der Teile von Datums- und Uhrzeitangaben enthalten sind aus.

Auswählen von Variablen zur Zusammenführung in eine einzige Datums-/Zeitvariable

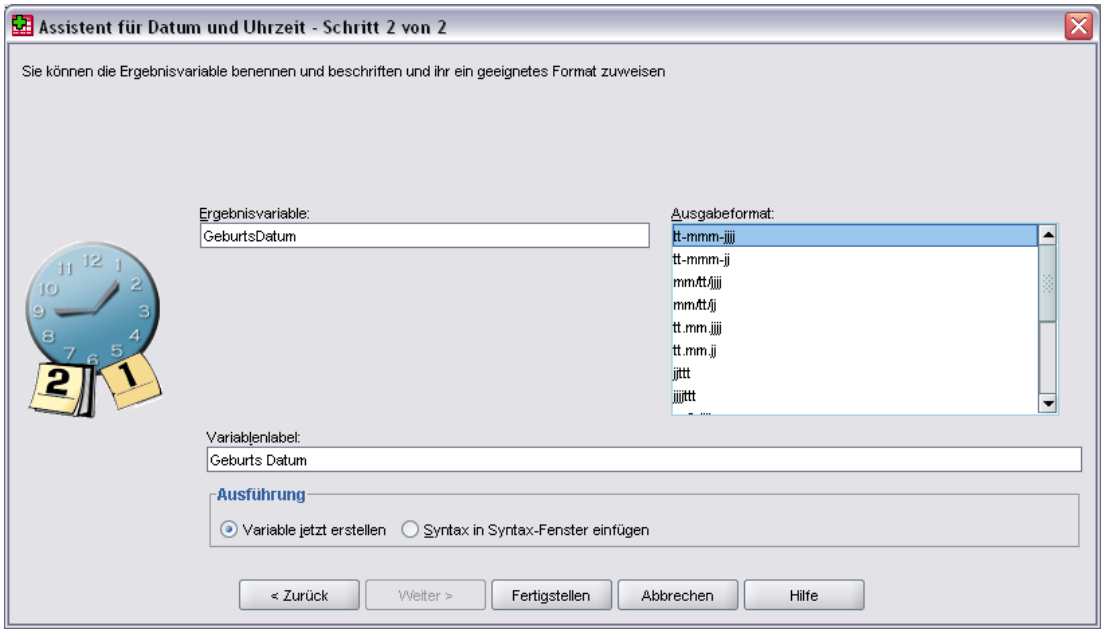
Abbildung 8-19
Erstellen einer Datums-/Zeitvariablen aus einem Variablen-Set, Schritt 1



- ▶ Wählen Sie die Variablen aus, die für die verschiedenen Teile des Datums bzw. der Uhrzeit stehen.
  - Einige Kombinationen sind nicht zulässig. So kann beispielsweise keine Datums-/Zeitvariable aus “Jahr” und “Tag im Monat” erstellt werden, da nach der Auswahl von “Jahr” ein vollständiges Datum erforderlich ist.
  - Bereits bestehende Datums-/Zeitvariablen können nicht als Teil der zu erstellenden Datums-/Zeitvariablen verwendet werden. Bei den Variablen für die Teile der neuen Datums-/Zeitvariablen muss es sich um ganze Zahlen handeln. Eine Ausnahme ist die zulässige Verwendung einer bestehenden Datums-/Zeitvariablen als Sekunden-Teil der neuen Variablen. Da Sekundenbruchteile zulässig sind, muss die für die Sekunden verwendete Variable keine ganze Zahl sein.
  - Werte für einen Teil der neuen Variablen, die nicht innerhalb des zulässigen Bereichs liegen, führen zum Wert “systemdefiniert fehlend” für die neue Variable. Wenn Sie beispielsweise für den Monat versehentlich eine Variable verwenden, die den Tag im Monat angibt, dann wird allen Fällen, bei denen der Wert für den Tag im Monat im Bereich von 14–31 liegt, der Wert “systemdefiniert fehlend” für die neue Variable zugewiesen, da der gültige Bereich für Monate in SPSS 1–13 beträgt.

Angeben der durch das Zusammenführen von Variablen erstellten Datums-/Zeitvariablen

Abbildung 8-20
Erstellen einer Datums-/Zeitvariablen aus einem Variablen-Set, Schritt 2



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.
- Wählen Sie ein Datums-/Uhrzeitformat aus der Liste “Ausgabeformat” aus.

  Die folgenden Optionen sind verfügbar:

  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

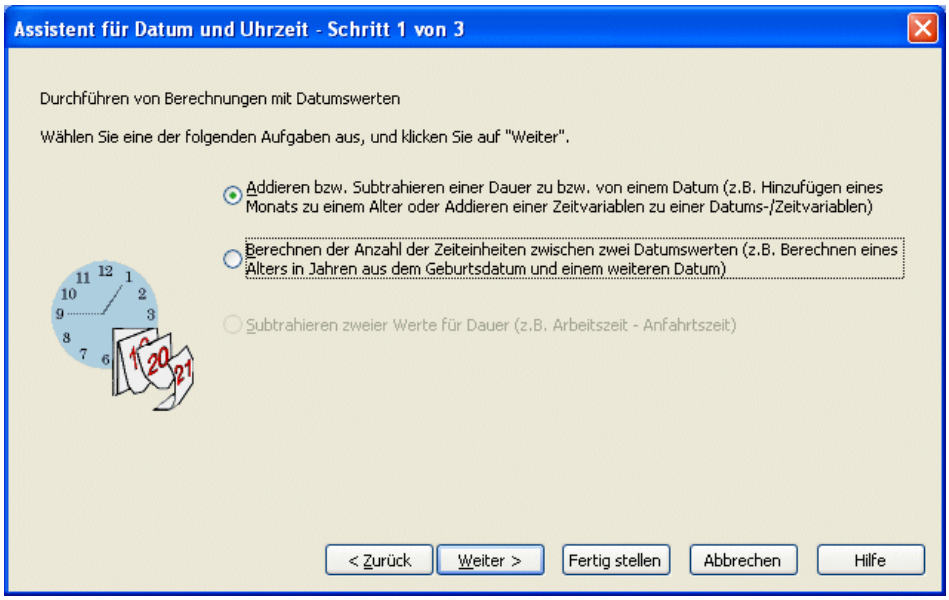
Addieren oder Subtrahieren von Werten zu bzw. von Datums-/Zeitvariablen

So können Sie Werte zu bzw. von Datums-/Zeitvariablen addieren bzw. subtrahieren:

- Wählen Sie im Einführungsbildschirm des Assistenten für Datum und Uhrzeit die Option Berechnungen mit Datums- und Zeitwerten durchführen aus.

Auswählen des Typs der mit Datums-/Zeitvariablen durchzuführenden Berechnung

Abbildung 8-21
Addieren oder Subtrahieren von Werten zu bzw. von Datums-/Zeitvariablen, Schritt 1



- **Addieren bzw. Subtrahieren einer Dauer zu bzw. von einem Datum.** Diese Option dient zum Addieren oder Subtrahieren von Werten zu bzw. von Variablen mit Datumsformat. Sie können eine Zeitdauer, die einen festen Wert aufweist (z. B. 10 Tage), oder die Werte aus einer numerischen Variablen (beispielsweise einer Variablen, die Jahre angibt) addieren bzw. subtrahieren.
- **Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten.** Mit dieser Option erhalten Sie die Differenz zwischen zwei Datumswerten (angegeben in einer von Ihnen ausgewählten Einheit). Beispielsweise können Sie die Anzahl der Jahre oder der Tage ermitteln, die zwischen zwei Datumsangaben liegt.
- **Subtrahieren zweier Werte für Dauer.** Mit dieser Option erhalten Sie die Differenz zwischen zwei Variablen, die ein Format für die Zeitdauer aufweisen, beispielsweise hh:mm oder hh:mm:ss.

Anmerkung: Aufgaben werden deaktiviert, wenn das Daten-Set nicht die für die Ausführung der Aufgabe erforderlichen Variablen aufweist. Beispiel: Wenn das Daten-Set nicht zwei Variablen mit einem Format für Zeitdauern aufweist, findet die Aufgabe zur Subtraktion zweier Werte für Dauer keine Anwendung und ist deaktiviert.

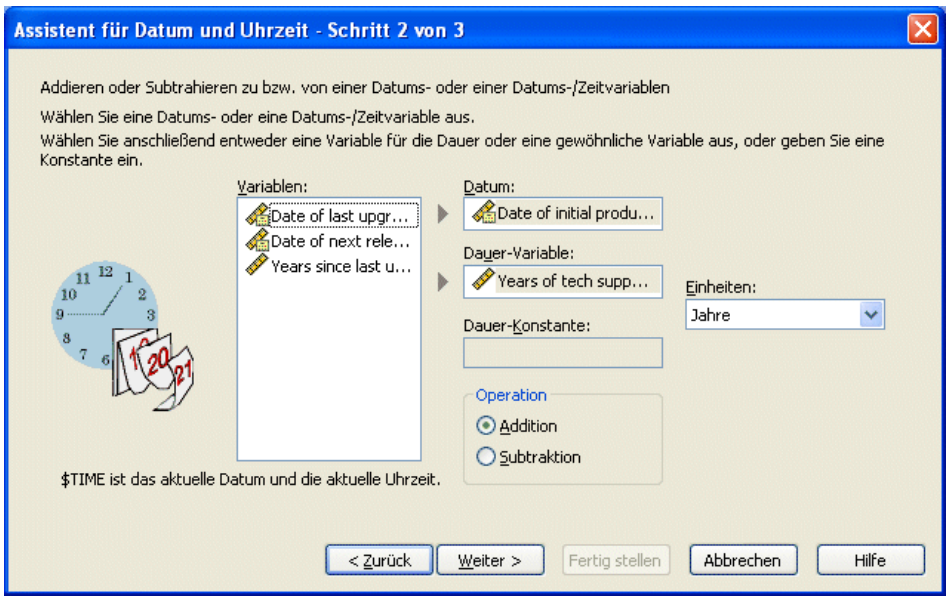
Addieren bzw. Subtrahieren einer Dauer zu bzw. von einem Datum

So addieren bzw. subtrahieren Sie eine Dauer zu bzw. von einer Variablen mit Datumsformat:

▶ Wählen Sie im Bildschirm *Durchführen von Berechnungen mit Datumswerten* des Assistenten für Datum und Uhrzeit die Option Addieren bzw. Subtrahieren einer Dauer zu bzw. von einem Datum aus.

Auswahl der Datums-/Zeitvariablen und der zu addierenden bzw. subtrahierenden Dauer

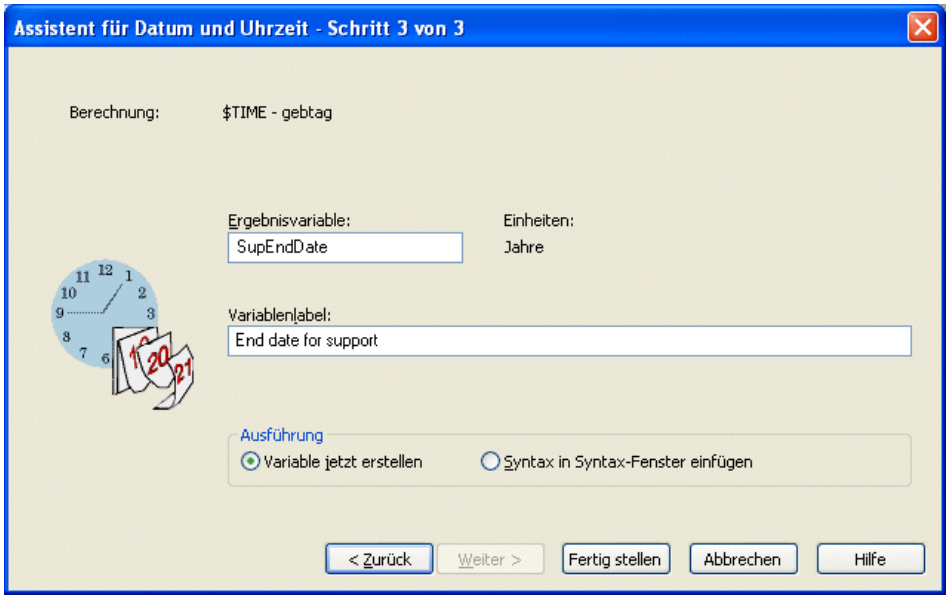
Abbildung 8-22
Addieren bzw. Subtrahieren einer Dauer, Schritt 2



- Wählen Sie eine Datums- oder Zeitvariable aus.
- Wählen Sie eine Dauer-Variable aus oder geben Sie einen Wert für die Dauer-Konstante ein. Variablen, die für die Dauer verwendet werden, können keine Datums- bzw. Datums-/Zeitvariablen sein. Es kann sich bei ihnen um Dauer-Variablen oder einfache numerische Variablen handeln.
- Wählen Sie die Einheit für die Dauer aus der Dropdown-Liste aus. Wählen Sie Dauer aus, wenn Sie eine Variable verwenden und die Variable in einem Format für die Dauer verwenden, beispielsweise hh:mm oder hh:mm:ss.

Angeben des Ergebnisses der Addition bzw. Subtraktion einer Dauer zu bzw. von einer Datums-/Zeitvariablen

Abbildung 8-23
Addieren bzw. Subtrahieren einer Dauer, Schritt 3



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.

  Die folgenden Optionen sind verfügbar:

  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

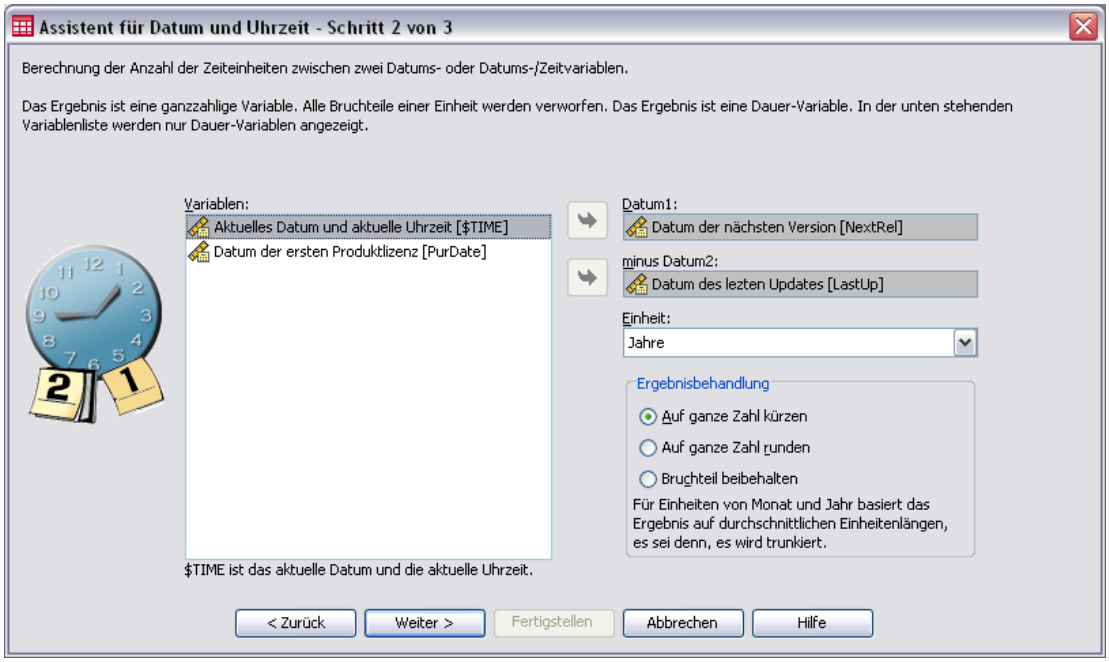
Subtrahieren von Variablen mit Datumsformat

So können Sie zwei Variablen mit Datumsformat subtrahieren:

- Wählen Sie im Bildschirm *Durchführen von Berechnungen mit Datumswerten* des Assistenten für Datum und Uhrzeit die Option Berechnen der Anzahl der Zeiteinheiten zwischen zwei Datumswerten aus.

Auswählen der zu subtrahierenden Variablen mit Datumsformat

Abbildung 8-24
Subtrahieren von Datumswerten, Schritt 2



- ▶ Wählen Sie die zu subtrahierenden Variablen aus.
- ▶ Wählen Sie die Einheit für das Ergebnis aus der Dropdown-Liste aus.
- ▶ Wählen Sie aus, wie das Ergebnis berechnet werden soll (Ergebnisbehandlung).

Ergebnisbehandlung.

Für die Berechnung des Ergebnisses stehen folgende Optionen zur Auswahl:

- **Auf ganze Zahl kürzen.** Alle Nachkommastellen des Ergebnisses werden ignoriert. Wenn beispielsweise das Datum 28.10.2006 vom Datum 21.10.2007 subtrahiert wird, lautet das Ergebnis 0 für die Jahre und 11 für die Monate.
- **Auf ganze Zahl runden.** Das Ergebnis wird auf die nächste ganze Zahl gerundet. Wenn beispielsweise das Datum 28.10.2006 vom Datum 21.10.2007 subtrahiert wird, lautet das Ergebnis 1 für die Jahre und 12 für die Monate.
- **Bruchteil beibehalten.** Der vollständige Wert wird beibehalten; es erfolgt keine Rundung oder Kürzung. Wenn beispielsweise das Datum 28.10.2006 vom Datum 21.10.2007 subtrahiert wird, lautet das Ergebnis 0,98 für die Jahre und 11,76 für die Monate.

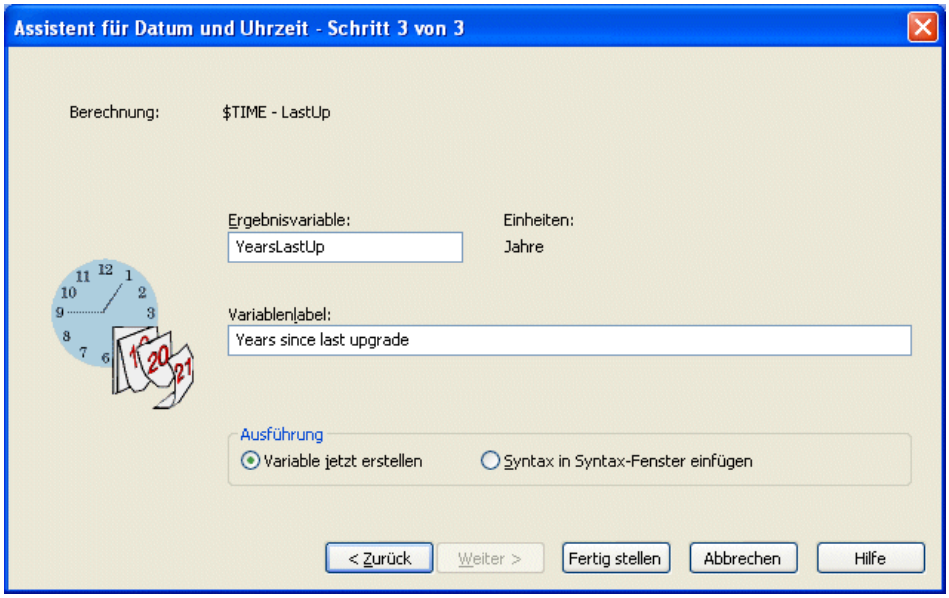
Beim Runden und Beibehalten des Bruchteils beruht das Ergebnis für die Jahre auf der durchschnittlichen Anzahl von Tagen im Jahr (365,25) und das Ergebnis für Monate auf der durchschnittlichen Anzahl von Tagen im Monat (30,3475). Wenn Sie beispielsweise im Format “m/t/j” den Wert 2/1/2007 von 3/1/2007 subtrahieren, ergibt sich (mit Dezimalstellen) ein Ergebnis von 0,92 Monaten; dagegen ergibt sich, wenn 3/1/2007 von 2/1/2007 subtrahiert wird,

eine Differenz von 1,02 Monaten. Dies wirkt sich auch auf Werte aus, die für Zeitspannen mit Schaltjahren berechnet werden. So ergibt sich, wenn 2/1/2008 von 3/1/2008 subtrahiert wird, unter Berücksichtigung der Dezimalstellen eine Differenz von 0,95 Monaten, im Gegensatz zum Wert 0,92 für dieselbe Zeitspanne ohne Schaltjahr.

Datum 1	Datum 2	Jahre			Monate		
		Kürzen	Runden	Bruch	Kürzen	Runden	Bruch
10/21/2006	10/28/2007	1	1	1.02	12	12	12.22
10/28/2006	10/21/2007	0	1	.98	11	12	11.76
2/1/2007	3/1/2007	0	0	.08	1	1	.92
2/1/2008	3/1/2008	0	0	.08	1	1	.95
3/1/2007	4/1/2007	0	0	.08	1	1	1.02
4/1/2007	5/1/2007	0	0	.08	1	1	.99

Angeben des Ergebnisses der Subtraktion von zwei Variablen mit Datumsformat

Abbildung 8-25
Subtrahieren von Datumswerten, Schritt 3



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.

Die folgenden Optionen sind verfügbar:

- Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

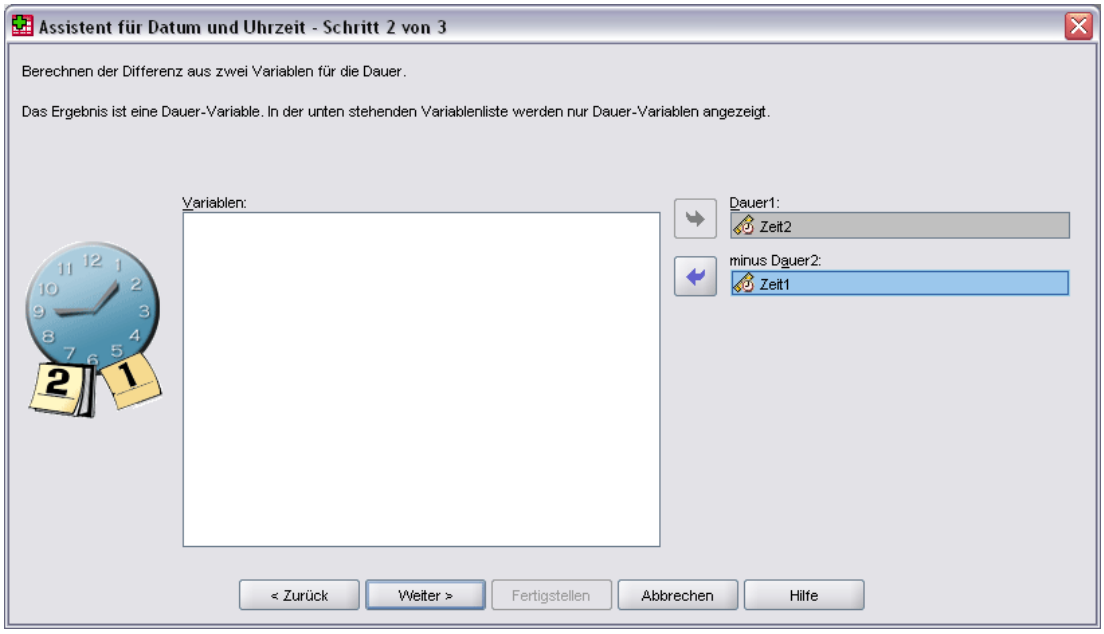
Subtrahieren von Dauer-Variablen

So können Sie zwei Dauer-Variablen subtrahieren:

- Wählen Sie im Bildschirm *Durchführen von Berechnungen mit Datumswerten* des Assistenten für Datum und Uhrzeit die Option Subtrahieren zweier Werte für Dauer aus.

Auswählen der zu subtrahierenden Dauer-Variablen

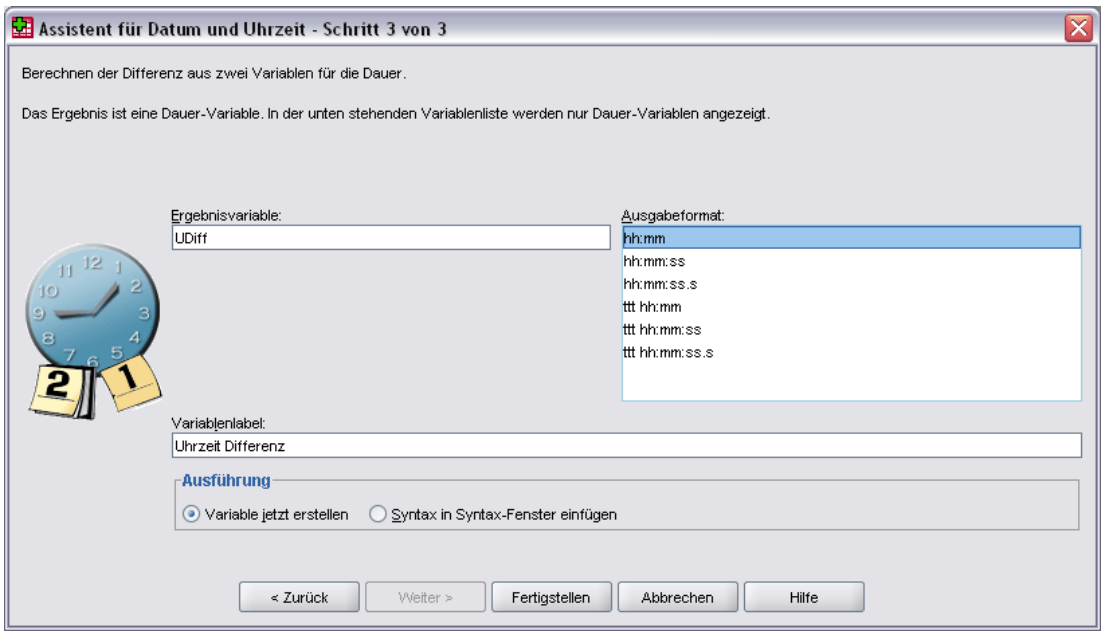
Abbildung 8-26
Subtrahieren zweier Werte für Dauer, Schritt 2



- Wählen Sie die zu subtrahierenden Variablen aus.

Angeben des Ergebnisses der Subtraktion von zwei Dauer-Variablen

Abbildung 8-27
Subtrahieren zweier Werte für Dauer, Schritt 3



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.
- Wählen Sie ein Format für die Dauer aus der Liste “Ausgabeformat” aus.

  Die folgenden Optionen sind verfügbar:

  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

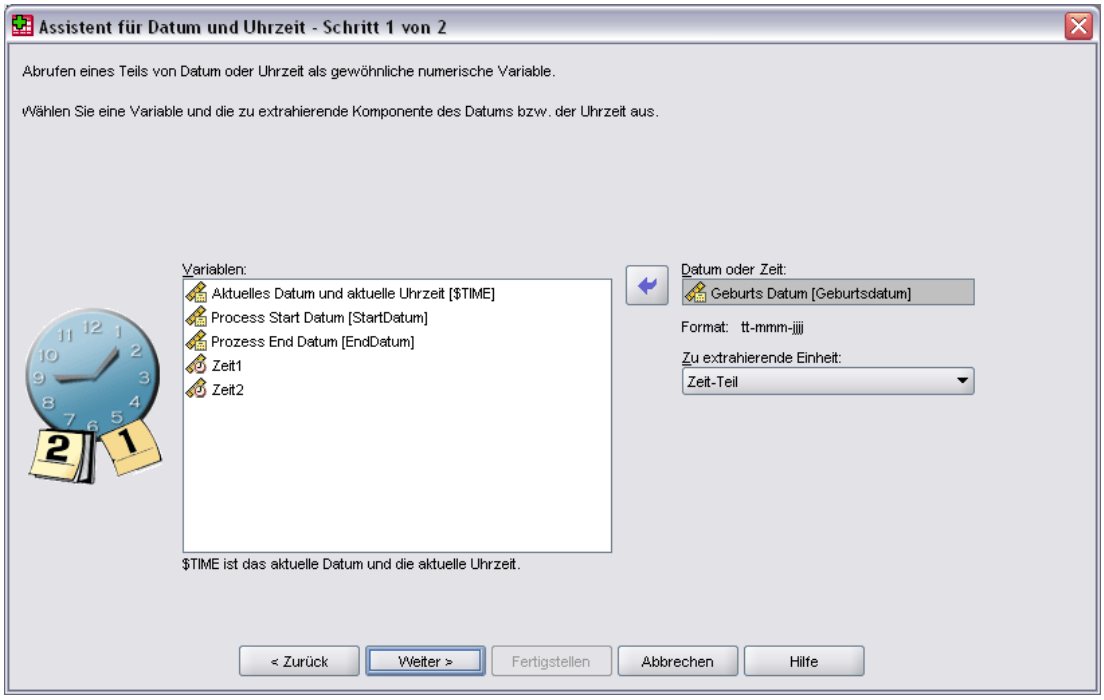
Extrahieren eines Teils einer Datums-/Zeitvariablen

So extrahieren Sie eine Komponente – beispielsweise das Jahr – aus einer Datums-/Zeitvariablen:

- Wählen Sie im Einführungsbildschirm des Assistenten für Datum und Uhrzeit die Option Einen Teil einer Datums- oder Zeitvariablen extrahieren aus.

Auswählen der aus der Datums-/Zeitvariablen zu extrahierenden Komponente

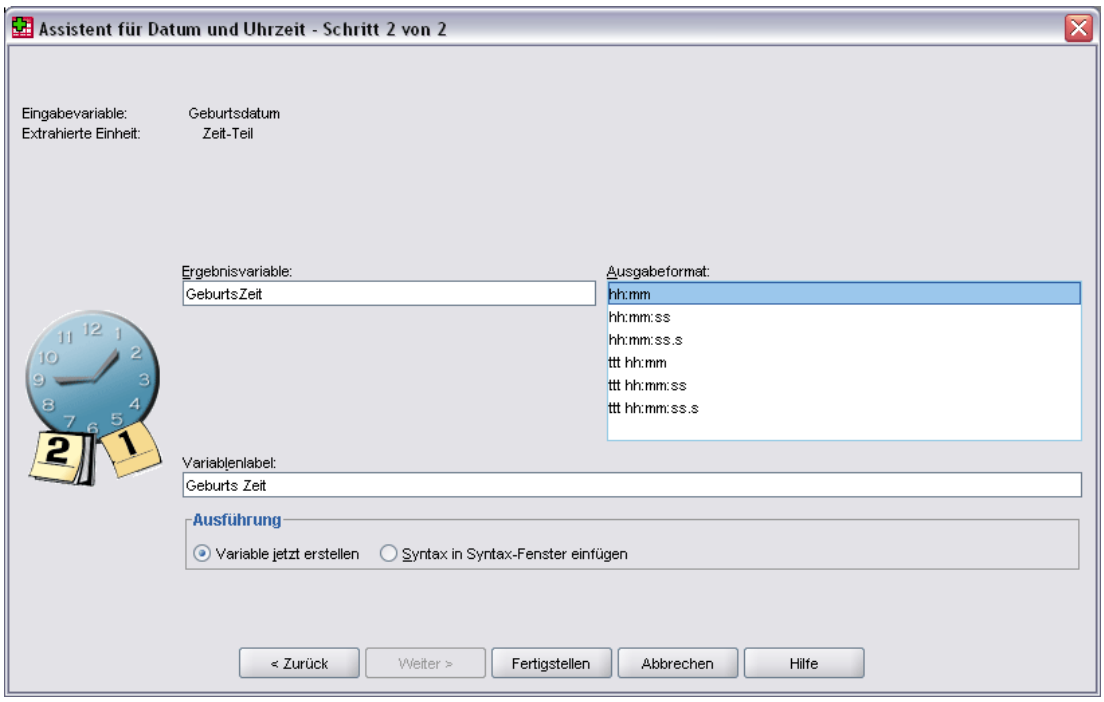
Abbildung 8-28
Abrufen eines Teils einer Datums-/Zeitvariablen, Schritt 1



- Wählen Sie die Variable aus, die den zu extrahierenden Datums- oder Zeitteil enthält.
- Wählen den zu extrahierenden Teil der Variablen aus der Dropdown-Liste aus. Sie können Informationen aus Datumsangaben extrahieren, die nicht explizit im angezeigten Datum enthalten sind, wie beispielsweise den Tag der Woche.

Angeben des Ergebnisses der Extraktion einer Komponente aus einer Datums-/Zeitvariablen

Abbildung 8-29
Abrufen eines Teils einer Datums-/Zeitvariablen, Schritt 2



- Geben Sie einen Namen für die Ergebnisvariable ein. Dieser darf nicht mit dem Namen einer bestehenden Variablen übereinstimmen.
- Beim Extrahieren des Datums- oder Zeitteils einer Datums-/Zeitvariablen müssen Sie in der Liste "Ausgabeformat" ein Format auswählen. Wenn kein Ausgabeformat erforderlich ist, ist die Liste "Ausgabeformat" deaktiviert.

  Die folgenden Optionen sind verfügbar:

  - Zuweisen eines beschreibenden Variablenlabels zur neuen Variablen.

Datentransformationen für Zeitreihen

In SPSS werden verschiedene Datentransformationen bereitgestellt, die bei Zeitreihenanalysen von Nutzen sind:

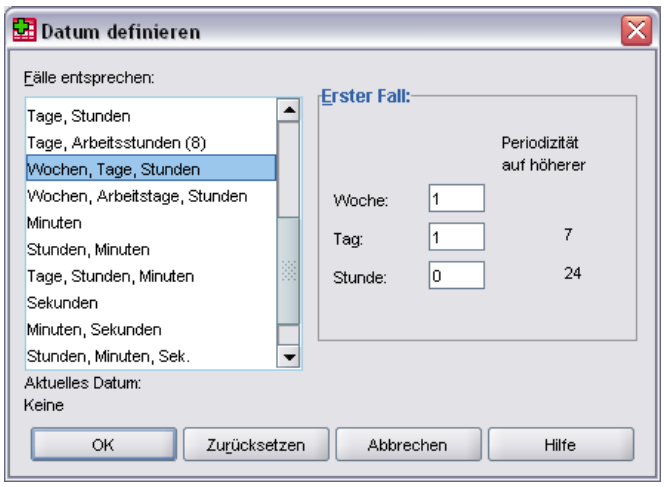
- Erzeugen von Datumsvariablen zum Erstellen von Periodizität und zum Unterscheiden zwischen historischen Perioden, Validierungsperioden und Vorhersageperioden,
- Erstellen von Zeitreihenvariablen als Funktionen vorhandener Zeitreihenvariablen,
- Ersetzen von system- und benutzerdefiniert fehlenden Werten durch Schätzwerte auf der Grundlage einer von mehreren möglichen Methoden.

Eine **Zeitreihe** wird erstellt, indem eine Variable (oder ein Variablen-Set) regelmäßig über einen Zeitraum beobachtet wird. Transformationen von Zeitreihendaten setzen eine Struktur in der Datendatei voraus, bei der jeder Fall (jede Zeile) eine Reihe von Beobachtungen zu einem unterschiedlichen Zeitpunkt darstellt und dabei die Zeitdauer zwischen den Fällen gleichförmig ist.

Datum definieren

Im Dialogfeld "Datum definieren" werden Datumsvariablen erstellt, die zum Herstellen der Periodizität einer **Zeitreihe** und zum Beschriften der Ausgabe aus Zeitreihenanalysen verwendet werden können.

Abbildung 8-30
Dialogfeld "Datum definieren"



Fälle entsprechen. Hiermit wird das zum Erstellen von Datumsangaben verwendete Zeitintervall definiert.

- Mit Kein Datum werden alle bisher definierten Datumsvariablen entfernt. Dabei werden Variablen mit den folgenden Namen gelöscht: *Jahr_*, *Quartal_*, *Monat_*, *Woche_*, *Tag_*, *Stunde_*, *Minute_*, *Sekunde_* und *Datum_*.
- Mit Benutzerdefiniert wird das Vorhandensein benutzerdefinierter Datumsvariablen angezeigt, die mit der Befehlssyntax erstellt wurden (z. B. eine viertägige Arbeitswoche). Dieser Eintrag spiegelt nur den aktuellen Stand der Arbeitsdatei wider. Die Auswahl dieses Eintrags in der Liste hat keine Auswirkung.

Erster Fall. Hiermit wird der Wert für das Startdatum definiert, das dem ersten Fall zugeordnet ist. Nachfolgenden Fällen werden auf dem Zeitintervall basierende sequentielle Werte zugeordnet.

Periodizität auf höherer Ebene. Hier wird die wiederholte zyklische Schwankung angezeigt, wie zum Beispiel die Anzahl der Monate in einem Jahr oder die Anzahl der Tage in einer Woche. Der angezeigte Wert ist der höchste Wert, den Sie eingeben können.

Für jede zum Definieren des Datums verwendete Komponente wird eine neue numerische Variable erzeugt. Die neuen Namen der Variablen enden mit einem Unterstrich. Es wird außerdem die beschreibende String-Variable *Datum_* aus den Komponenten erzeugt. Wenn Sie zum Beispiel

Wochen, Tage, Stunden ausgewählt haben, werden vier neue Variablen erstellt: *Woche_*, *Tag_*, *Stunde_* und *Datum_*.

Wenn bereits Datumsvariablen definiert wurden, werden diese beim Definieren neuer Datumsvariablen mit gleichem Namen ersetzt.

So definieren Sie Datumsangaben für Zeitreihendaten:

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Daten
  Datum definieren...
- Wählen Sie ein Zeitintervall aus der Liste "Fälle entsprechen:" aus.
- Geben Sie die Werte ein, die das Startdatum für "Erster Fall" definieren. Hiermit wird das dem ersten Fall zugewiesene Datum bestimmt.

Vergleich von Datumsvariablen und Variablen im Datumsformat

Mit "Datum definieren" erzeugte Datumsvariablen dürfen nicht mit Variablen im Datumsformat verwechselt werden, die in der Variablenansicht des Daten-Editors definiert werden. Datumsvariablen werden verwendet, um Periodizität für Zeitreihendaten zu erstellen. Variablen im Datumsformat stellen in verschiedenen Datums- und Uhrzeitformaten angezeigte Datums- und Uhrzeitangaben dar. Datumsvariablen sind einfache ganze Zahlen, welche die Anzahl von Tagen, Wochen, Stunden usw. ab einem benutzerdefinierten Ausgangspunkt angeben. Intern werden die meisten Variablen im Datumsformat als die Anzahl der seit dem 14. Oktober 1582 vergangenen Sekunden gespeichert.

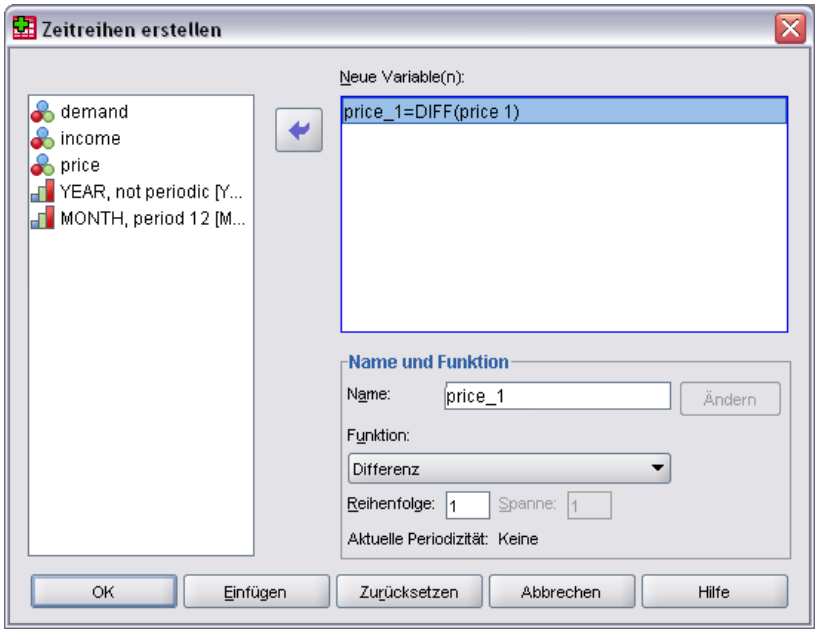
Zeitreihen erstellen

Im Dialogfeld "Zeitreihen erstellen" werden neue Variablen auf der Grundlage der Funktionen von vorhandenen numerischen **Zeitreihenvariablen** erstellt. Diese transformierten Werte werden in vielen Prozeduren zur Zeitreihenanalyse benutzt.

Neue Variablennamen bestehen in der Standardeinstellung aus den ersten sechs Zeichen der vorhandenen Variablen, aus denen sie erstellt wurden, einem Unterstrich und einer laufenden Nummer. Der neue Variablenname für die Variable *Preis* ist z. B. *Preis_1*. Den neuen Variablen werden alle definierten Wertelabels der ursprünglichen Variablen zugewiesen.

Zu den verfügbaren Funktionen zum Erzeugen von Zeitreihenvariablen gehören Differenzen, gleitende Durchschnitte, gleitende Mediane, Intervall- und Vorlauffunktionen.

Abbildung 8-31
Dialogfeld "Zeitreihen erstellen"



So erstellen Sie eine neue Zeitreihenvariable:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Transformieren
  Zeitreihen erstellen...

- Wählen Sie die Zeitreihenfunktion, die Sie zum Transformieren der ursprünglichen Variablen verwenden möchten.

- Wählen Sie die Variablen, aus denen Sie neue Zeitreihenvariablen erstellen möchten. Es können nur numerische Variablen verwendet werden.

  Die folgenden Optionen sind verfügbar:

  - Durch Eingabe neuer Variablennamen können die vorgegebenen Variablennamen überschrieben werden.
  - Ändern der Funktion für eine ausgewählte Variable.

Funktionen zur Transformation von Zeitreihen

Differenz. Die nichtsaisonale Differenz zwischen aufeinander folgenden Werten in den Datenreihen. Die Ordnung ist die Anzahl der zum Berechnen der Differenz verwendeten zurückliegenden Werte. Da für jede Ordnung der Differenzen eine Beobachtung fehlt, sind systemdefiniert fehlende Werte am Anfang der Datenreihe vorhanden. Wenn die Ordnung der Differenzen zum Beispiel 2 beträgt, besitzen die ersten zwei Fälle den systemdefiniert fehlenden Wert in der neuen Variablen.

Saisonale Differenz. Differenz zwischen Reihenwerten, die eine konstante Spanne auseinander liegen. Die Spanne basiert auf der aktuell definierten Periodizität. Zum Berechnen saisonaler Differenzen müssen Datumsvariablen (Menü "Daten", Befehl "Datum definieren") mit einer periodischen Komponente (wie den Monaten eines Jahres) definiert sein. Die Ordnung ist die Anzahl der zum Berechnen der Differenz verwendeten saisonalen Perioden. Die Anzahl der Fälle mit dem systemdefiniert fehlenden Wert am Anfang der Datenreihen ist gleich dem Produkt aus der Periodizität und der Ordnung. Wenn zum Beispiel die aktuelle Periodizität 12 und die Reihenfolge 2 beträgt, besitzen die ersten 24 Fälle den systemdefiniert fehlenden Wert als neue Variable.

Zentrierter gleitender Durchschnitt. Durchschnitt einer Spanne von Datenreihenwerten, die den aktuellen Wert umgeben und einschließen. Die Spanne ist die Anzahl der zum Berechnen des Durchschnitts verwendeten Datenreihenwerte. Wenn der Wert der Spanne gerade ist, wird der gleitende Durchschnitt so berechnet, dass für jedes Paar nichtzentrierter Mittelwerte der Durchschnitt gebildet wird. Die Anzahl der Fälle mit dem systemdefiniert fehlenden Wert am Anfang und Ende der Datenreihe für eine Spanne von *n* ist gleich *n*/2 bei geraden Werten für die Spanne und (*n*–1)/2 bei ungeraden Werten. Wenn die Spanne zum Beispiel 5 beträgt, gibt es 2 Fälle mit dem systemdefiniert fehlenden Wert am Anfang und am Ende der Datenreihe.

Zurückgreifender gleitender Durchschnitt. Durchschnitt der Spanne von Datenreihenwerten vor dem aktuellen Wert. Die Spanne ist die Anzahl der zum Berechnen des Durchschnitts verwendeten vorangehenden Datenreihenwerte. Die Anzahl der Fälle mit dem systemdefiniert fehlenden Wert am Anfang der Datenreihe ist gleich dem Wert der Spanne.

Gleitende Mediane. Median einer Spanne von Datenreihenwerten, die den aktuellen Wert umgeben und einschließen. Die Spanne ist die Anzahl der zum Berechnen des Medians verwendeten Datenreihenwerte. Wenn die Spanne geradzahlig ist, wird der Median durch Berechnen des Durchschnitts jedes Paars unzentrierter Mediane ermittelt. Die Anzahl der Fälle mit dem systemdefiniert fehlenden Wert am Anfang und Ende der Datenreihe für eine Spanne von *n* ist gleich *n*/2 bei geraden Werten für die Spanne und (*n*–1)/2 bei ungeraden Werten. Wenn die Spanne zum Beispiel 5 beträgt, gibt es 2 Fälle mit dem systemdefiniert fehlenden Wert am Anfang und am Ende der Datenreihe.

Kumulierte Summe. Kumulierte Summe der Datenreihenwerte bis zum und einschließlich des aktuellen Werts.

Lag. Wert eines zurückliegenden Falls auf der Grundlage der angegebenen Ordnung der Intervalle. Die Ordnung ist die Anzahl der Fälle vor dem gegenwärtigen Fall, aus dem der Wert ermittelt wird. Die Anzahl der Fälle mit dem systemdefiniert fehlenden Wert am Anfang der Datenreihe ist gleich der Ordnung.

Lead. Wert eines nachfolgenden Falls auf der Grundlage der angegebenen Ordnung der Intervalle. Die Ordnung ist die Anzahl der Fälle nach dem aktuellen Fall, aus dem der Wert ermittelt wird. Die Anzahl der Fälle mit dem systemdefiniert fehlenden Wert am Ende der Datenreihe ist gleich der Ordnung.

Glätten. Neue Datenreihenwerte auf der Grundlage einer Glättung von zusammengesetzten Daten. Die Glättung beginnt mit einem gleitenden Median von 4, der von einem gleitenden Median von 2 zentriert wird. Dann werden die Werte erneut durch Anwendung eines gleitenden Medians von 5, eines gleitenden Medians von 3 und Hanning (gleitende gewichtete Mittelwerte) geglättet. Residuen werden durch Subtrahieren der geglätteten Datenreihen von den ursprünglichen Datenreihen berechnet. Dieser vollständige Prozess wird dann erneut auf die errechneten

Residuen angewendet. Zuletzt werden durch Subtrahieren der beim ersten Durchlauf dieses Prozesses errechneten Werte die geglätteten Residuen ermittelt. Dieses Verfahren wird auch als **T4253H-Glättung** bezeichnet.

Fehlende Werte ersetzen

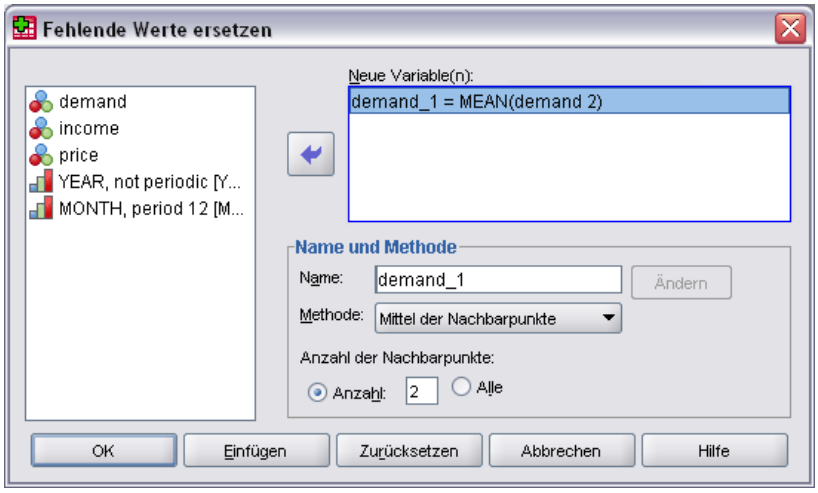
Fehlende Beobachtungen können Probleme in der Analyse aufwerfen, und einige Maße für Zeitreihen können bei fehlenden Werten in den Datenreihen nicht berechnet werden. In einigen Fällen ist der Wert für eine bestimmte Beobachtung einfach nicht bekannt. Fehlende Daten können auch aus den folgenden Ursachen entstehen:

- Jeder Grad der Differenzierung verkürzt eine Reihe um ein Element.
- Jeder Grad der saisonalen Differenzierung verkürzt eine Reihe um eine Saison.
- Wenn Sie eine neue Reihe mit Prognosen erstellen, die über das Ende der vorhandenen Reihe hinausreichen (indem Sie auf die Schaltfläche Speichern klicken und eine geeignete Auswahl treffen), weisen die ursprüngliche Reihe und die erzeugte Residuenreihe fehlende Werte für die neuen Beobachtungen auf.
- Einige Transformationen (z. B. die Log-Transformation) erzeugen fehlende Daten für bestimmte Werte in der ursprünglichen Reihe.

Fehlende Daten am Anfang oder Ende einer Zeitreihe stellen kein größeres Problem dar. Sie verkürzen nur die brauchbare Länge der Zeitreihe. Lücken im Inneren einer Zeitreihe (*eingebettete* fehlende Daten) können ein viel schwerwiegenderes Problem darstellen. Das Ausmaß des Problems ist abhängig vom eingesetzten Analyseverfahren.

Im Dialogfeld “Fehlende Werte ersetzen” werden neue **Zeitreihenvariablen** aus bereits vorhandenen erstellt. Dabei werden fehlende Werte durch Schätzwerte ersetzt, die mit einer von mehreren möglichen Methoden errechnet werden. Neue Variablennamen bestehen in der Standardeinstellung aus den ersten sechs Zeichen der vorhandenen Variablen, aus denen sie erstellt wurden, einem Unterstrich und einer laufenden Nummer. Der neue Variablenname für die Variable *Preis* ist z. B. *Preis_1*. Den neuen Variablen werden alle definierten Wertelabels der ursprünglichen Variablen zugewiesen.

Abbildung 8-32
Dialogfeld "Fehlende Werte ersetzen"



So ersetzen Sie fehlende Werte für Zeitreihenvariablen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
Fehlende Werte ersetzen...

▶ Wählen Sie die zum Ersetzen fehlender Werte zu verwendende Schätzmethode aus.

▶ Wählen Sie die Variablen, für die Sie fehlende Werte ersetzen möchten.

Die folgenden Optionen sind verfügbar:

- Durch Eingabe neuer Variablennamen können die vorgegebenen Variablennamen überschrieben werden.
- Die Schätzmethode für eine ausgewählte Variable kann geändert werden.

Schätzmethoden zum Ersetzen fehlender Werte

Zeitreihen-Mittelwert. Fehlende Werte werden durch den Mittelwert der gesamten Datenreihe ersetzt.

Mittel der Nachbarpunkte. Fehlende Werte werden durch den Mittelwert gültiger Umgebungswerte ersetzt. Die Anzahl der Nachbarpunkte ist die Anzahl gültiger Werte über und unter dem zum Berechnen des Mittelwerts verwendeten fehlenden Wert.

Median der Nachbarpunkte. Fehlende Werte werden mit dem Median gültiger Umgebungswerte ersetzt. Die Anzahl der Nachbarpunkte ist die Anzahl gültiger Werte über und unter dem zum Berechnen des Medians verwendeten fehlenden Wert.

Lineare Interpolation. Fehlende Werte werden unter Anwendung einer linearen Interpolation ersetzt. Für die Interpolation werden der letzte gültige Wert vor dem fehlenden Wert und der erste gültige Wert nach dem fehlenden Wert verwendet. Wenn der erste oder letzte Fall in der Datenreihe einen fehlenden Wert hat, wird der fehlende Wert nicht ersetzt.

Linearer Trend am Punkt. Fehlende Werte werden mit dem linearen Trend für den Punkt ersetzt. Für die vorhandene Datenreihe wird eine Regression auf eine von 1 bis *n* skalierte Indexvariable ausgeführt. Fehlende Werte werden mit ihren vorhergesagten Werten ersetzt.

Bewerten von Daten mit Vorhersagemodellen

Die Anwendung eines Vorhersagemodells auf eine Datenmenge wird als **Bewertung** (Scoring) der Daten bezeichnet. SPSS, Clementine und AnswerTree bieten Verfahren für den Aufbau von Vorhersagemodellen, wie Regression, Clustern, sowie Baummodelle und Modelle für neurale Netzwerke. Sobald ein Modell erstellt wurde, können die Modellspezifikationen als XML-Datei mit allen für die Rekonstruktion des Modells erforderlichen Informationen gespeichert werden. Das SPSS Server-Produkt bietet anschließend die Möglichkeit, eine XML-Modelldatei zu lesen und das Modell auf ein Daten-Set anzuwenden.

Beispiel. Eine Kreditanwendung wird auf der Grundlage verschiedener Aspekte des Bewerbers und des betreffenden Kredits nach Risiko bewertet. Anhand des aus dem Risikomodell gewonnenen Score-Werts für den Kredit wird die Kreditanwendung angenommen oder zurückgewiesen.

Die Bewertung (Scoring) wird als Datentransformation behandelt. Das Modell wird intern als Menge numerischer Transformationen ausgedrückt, die auf ein vorgegebenes Variablen-Set (die im Modell festgelegten Einflussvariablen) angewendet werden, um ein vorhergesagtes Ergebnis zu erzielen. In diesem Sinne ist der Vorgang der Bewertung von Daten mithilfe eines bestimmten Modells im Grunde dasselbe wie die Anwendung jeder anderen Funktion, wie beispielsweise der Funktion “Quadratwurzel”, auf eine Datenmenge.

Die Bewertung ist nur mit SPSS Server verfügbar und kann von Benutzern interaktiv ausgeführt werden, die im Modus für verteilte Analysen arbeiten. Beim Scoren von großen Datendateien ist die Verwendung von SPSS Batch Facility zu empfehlen. Dies ist eine gesonderte, ausführbare Datei, die im Lieferumfang von SPSS Server enthalten ist. Informationen zur Verwendung von SPSS Batch Facility finden Sie imBenutzerhandbuch zu SPSS Batch Facility, das Sie als PDF-Datei auf der Produkt-CD von SPSS Server finden.

Der Bewertungsprozess besteht aus den folgenden Schritten:

- Laden eines Modells aus einer Datei im XML- (PMML-)Format.
- Berechnen der Scores als neue Variable unter Verwendung der Funktion ApplyModel oder StrApplyModel im Dialogfeld Variable berechnen.

Weitere Informationen über die Funktionen `ApplyModel` und `StrApplyModel` finden Sie im Handbuch *Command Syntax Reference* im Abschnitt “Transformation Expressions” unter “Scoring Expressions”.

In der folgenden Tabelle werden die Prozeduren aufgeführt, die den Export von Modellspezifikationen in XML unterstützen. Die exportierten Modelle können mit SPPS Server zur Bewertung neuer Daten verwendet werden, wie oben beschrieben. Die vollständige Liste

der Modelltypen, die mit SPSS Server gescort werden können, finden Sie in der Beschreibung der Funktion `ApplyModel`.

Name der Prozedur	Name des Befehls	Option
Diskriminanz	`DISCRIMINANT`	Base
Lineare Regression	`REGRESSION`	Base
Two-Step-Clusteranalyse	`TWOSTEP CLUSTER`	Base
Verallgemeinerte lineare Modelle	`GENLIN`	Advanced Models
Allgemeines lineares Modell für komplexe Stichproben	`CSGLM`	Komplexe Stichproben
Logistische Regression für komplexe Stichproben	`CSLOGISTIC`	Komplexe Stichproben
Ordinale Regression für komplexe Stichproben	`CSORDINAL`	Komplexe Stichproben
Logistische Regression	`LOGISTIC REGRESSION`	Regression
Multinomiale logistische Regression	`NOMREG`	Regression
Klassifizierungsbaum	`TREE`	Tree

Laden eines gespeicherten Modells

Mithilfe des Dialogfelds “Modell laden” können Sie Vorhersagemodelle laden, die im XML-(PMML-)Format gespeichert wurden. Das Dialogfeld ist nur verfügbar, wenn Sie im Modus für verteilte Analysen arbeiten. Sie müssen ein Modell laden, ehe Sie damit Daten scoren können.

Abbildung 8-33
Ausgabe beim Laden eines Modells

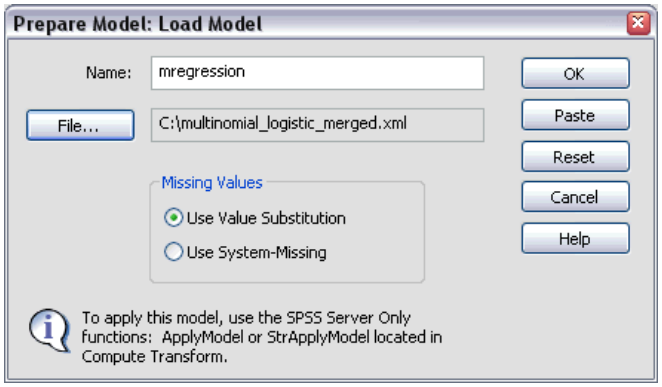
	Model Variables					
SPSS Variables	Name	Label	Type	Width	Role	Measurement
Age_Group	Age_Group	Age group	Numeric	8	Factor	Nominal
Frequency	Frequency	Number of purchases	Numeric	7	Covariate	Continuous
Gender	Gender	Gender	String	6	Factor	Nominal
Has_Broadband	Has_Broadband	Has home broadband connection	Numeric	10	Factor	Nominal
Has_Child	Has_Child	Has children	Numeric	10	Factor	Nominal
Income_Group	Income_Group	Income group	Numeric	10	Factor	Nominal
Log_Amount	Log_Amount	Log-transformed amount	Numeric	8	Covariate	Continuous
Recency	Recency	Days since last purchase	Numeric	10	Covariate	Continuous
Response	Response	Response	Numeric	10	Target	Nominal

So laden Sie ein Modell:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

Transformieren
  Modell vorbereiten
    Modell laden...

Abbildung 8-34
Dialogfeld "Prepare Model: Load Model"



- Geben Sie einen Namen ein, der diesem Modell zugeordnet werden soll. Jedes geladenen Modell muss einen eindeutigen Namen aufweisen.

- Klicken Sie auf Datei und wählen Sie eine Modelldatei aus. Im dadurch geöffneten Dialogfeld "Datei öffnen" werden die Dateien angezeigt, die im Modus für verteilte Analysen verfügbar sind. Hierbei handelt es sich um die Dateien auf dem Computer, auf dem der SPSS Server installiert wurde, und um Dateien in freigegebenen Ordnen oder Laufwerken auf Ihrem lokalen Computer.

  Anmerkung: Wenn Sie Daten bewerten, wird das Modell auf Variablen in der Arbeitsdatei mit denselben Namen wie die Variablen aus der Modelldatei angewendet. Sie können Variablen aus dem ursprünglichen Modell anderen Variablen in der Arbeitsdatei zuordnen, indem Sie die Befehlssyntax verwenden (Informationen hierzu finden unter dem Befehl `MODEL HANDLE`).

  Name. Ein Name zum Identifizieren dieses Modells. Es gelten dieselben Regeln für gültige Modellnamen wie für Variablennamen (siehe Variablennamen in Kapitel 5 auf S. 83), wobei zusätzlich das Zeichen "$" als erstes Zeichen zulässig ist. Sie verwenden diesen Namen, um das Modell anzugeben, wenn Sie die Daten mit den Funktionen ApplyModel oder StrApplyModel scoren.

  Datei. Die XML- (PMML-) Datei, die die Modellspezifikation enthält.

Fehlende Werte

In diesem Gruppenfeld legen Sie die Behandlung von fehlenden Werten für die im Modell definierten Einflussvariablen fest, die während des Bewertungsprozesses auftreten. Bei der Bewertung liegt in den folgenden Fällen ein fehlender Wert vor:

- Eine Einflussvariable enthält keinen Wert. Bei numerischen Variablen entspricht dies dem systemdefinierten fehlenden Wert. Bei String-Variablen entspricht dies einem leeren String.
- Der Wert für die vorliegende Einflussvariable wurde im Modell als benutzerdefiniert fehlend definiert. Werte, die in der Arbeitsdatei, aber nicht im Modell, als benutzerdefiniert fehlend definiert sind, werden im Bewertungsprozess nicht als fehlende Werte behandelt.
- Die Einflussvariable ist kategorial und der Wert entspricht keiner der im Modell definierten Kategorien.

Werte ersetzen. Hierbei wird versucht, fehlende Werte beim Bewerten von Fällen zu ersetzen. Die Methode für das Bestimmen eines Werts, der einen fehlenden Wert ersetzen soll, hängt von der Art des Vorhersagemodells ab.

- **SPSS Modelle.** Wenn beim Erstellen und Speichern eines linearen Regressionsmodells oder eines Diskriminanzmodells für die unabhängigen Variablen festgelegt wurde, dass fehlende Werte durch den Mittelwert ersetzt werden sollen, wird bei der Berechnung der Bewertung dieser Mittelwert anstelle des fehlenden Werts verwendet. Wenn der Mittelwert nicht verfügbar ist, wird der systemdefinierte fehlende Wert zurückgegeben.
- **AnswerTree-Modelle und Modelle des Befehls TREE.** Bei CHAID- und Exhaustive CHAID-Modellen wird für eine fehlende Teilungsvariable der größte untergeordnete Knoten ausgewählt. Der größte untergeordnete Knoten ist der Knoten mit der größten Grundgesamtheit unter den untergeordneten Knoten auf der Grundlage einer Stichprobe von Lernfällen. Für C&RT- und QUEST-Modelle werden (wenn überhaupt) zuerst Ersatzteilungsvariablen verwendet. (Ersatzteilungsvariablen sind Teilungsvariablen, mit denen versucht wird, anhand von anderen Einflussvariablen eine möglichst starke Übereinstimmung mit der ursprünglichen Teilung zu erzielen.) Wenn keine Ersatzteilungsvariablen angegeben werden oder alle Ersatzteilungsvariablen fehlen, wird der größte untergeordnete Knoten verwendet.
- **Clementine-Modelle.** Lineare Regressionsmodelle werden behandelt wie für SPSS-Modelle erläutert. Logistische Regressionsmodelle werden behandelt, wie unter "Logistische Regressionsmodelle" beschrieben. C&RT Tree-Modelle werden behandelt, wie im Abschnitt zu C&RT-Modellen unter "AnswerTree-Modelle" beschrieben.
- **Logistische Regressionsmodelle.** Wenn bei Kovariaten in logistischen Regressionsmodellen ein Mittelwert der Einflussvariablen als Teil des gespeicherten Modells aufgenommen wurde, wird bei der Berechnung der Bewertung dieser Mittelwert anstelle des fehlenden Werts verwendet. Wenn die Einflussvariable kategorial ist (z. B. ein Faktor in einem logistischen Regressionsmodell) oder wenn der Mittelwert nicht verfügbar ist, wird der systemdefinierte fehlende Wert zurückgegeben.

Systemdefinierte fehlende Werte verwenden. Beim Bewerten eines Falls mit einem fehlenden Wert wird der systemdefinierte fehlende Wert zurückgegeben.

Anzeigen einer Liste der geladenen Modelle

Sie können eine Liste der aktuell geladenen Modelle erhalten. Wählen Sie die folgenden Befehle aus den Menüs aus (nur im Modus für verteilte Analysen verfügbar):

Transformieren
  Modell vorbereiten
    Modell(e) auflisten

Hiermit wird eine Tabelle mit den Modell-Bezeichnungen erstellt. Die Tabelle enthält eine Liste aller aktuell geladenen Modelle sowie für jedes Modell den zugeordneten Namen (die so genannten Modell-Bezeichnung), den Modelltyp, den Pfad zur Modelldatei und die Methode zur Behandlung fehlender Werte.

Abbildung 8-35
Liste der geladenen Modelle

Model Handles

Name	Type	Input File Name	Validation handling	Missing handling
mregression	NOMREG	C:\models\mregression.xml	ON	SUBSTITUTE
tree	TREE	C:\models\tree.xml	ON	SUBSTITUTE

Zusätzliche Funktionen bei der Befehlssyntax

Sie können die ausgewählten Optionen im Dialogfeld “Modell laden” in ein Syntaxfenster einfügen und die entsprechende Befehlssyntax von `MODEL HANDLE` bearbeiten. Hierbei haben Sie folgende Möglichkeiten:

- Sie können Variablen aus dem ursprünglichen Modell anderen Variablen in der Arbeitsdatei zuordnen (mit dem Unterbefehl `MAP`). In der Standardeinstellung wird das Modell auf Variablen in der Arbeitsdatei mit denselben Namen wie die Variablen aus der Modelldatei angewendet.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
9

Umgang mit Dateien und Dateitransformationen

Datendateien liegen nicht immer genau in der Form vor, die Sie gerade benötigen. Vielleicht möchten Sie Datendateien zusammenfügen, die Daten in einer anderen Reihenfolge sortieren lassen, eine Teilmenge von Fällen auswählen oder die Einheit für die Analyse durch Gruppieren von Fällen ändern. In SPSS stehen Ihnen mehrere Möglichkeiten zum Transponieren von Dateien zur Verfügung, darunter folgende Funktionen:

Sortieren von Daten. Sie können Fälle nach dem Wert einer oder mehrerer Variablen sortieren lassen.

Transponieren von Fällen und Variablen. Das Format für SPSS-Datendateien ist so definiert, dass Zeilen als Fälle und Spalten als Variablen eingelesen werden. Bei Datendateien, in denen diese Reihenfolge umgekehrt ist, können Sie die Zeilen und Spalten vertauschen und so die Daten im richtigen Format einlesen.

Zusammenfügen von Dateien. Sie können zwei oder mehr Datendateien zusammenfügen. Es können Dateien zusammengefügt werden, welche dieselben Variablen, aber verschiedene Fälle enthalten, oder Dateien mit denselben Fällen und unterschiedlichen Variablen.

Auswählen von Teilmengen von Fällen. Sie können die Analyse auf eine Teilmenge von Fällen beschränken oder Analysen für verschiedene Teilmengen gleichzeitig vornehmen.

Aggregieren von Daten. Sie können die Einheit für die Analyse ändern, indem Sie Fälle nach Wert(en) einer oder mehrerer Gruppenvariablen aggregieren.

Gewichten von Daten. Sie können Fälle für die Analyse nach dem Wert einer Gewichtungsvariablen gewichten.

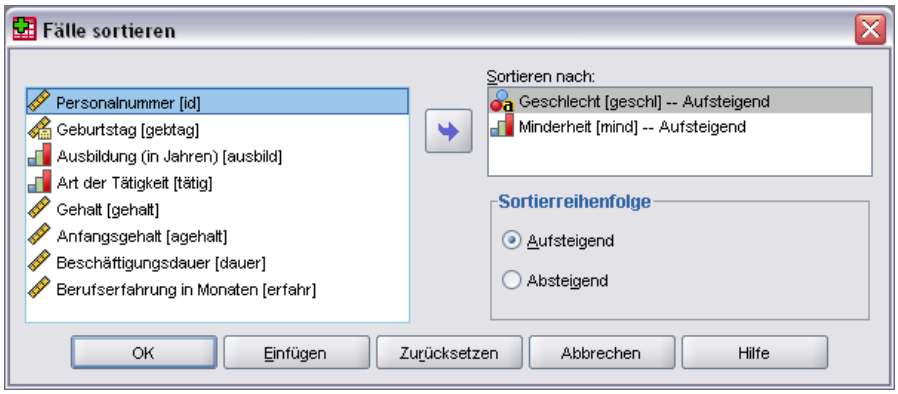
Umstrukturieren von Daten. Sie können Daten umstrukturieren und somit einen Fall (Datensatz) aus mehreren Fällen bzw. mehrere Fälle aus einem Fall erstellen.

Fälle sortieren

Mit diesem Dialogfeld können Sie Fälle (also Zeilen) der Datendatei nach einer oder mehreren Sortiervariablen sortieren. Sie können Fälle in aufsteigender oder absteigender Folge sortieren.

- Wenn Sie mehrere Sortiervariablen auswählen, werden die Fälle nach Variablen in den Kategorien der vorhergehenden Variablen aus der Liste “Sortieren nach” sortiert. Wenn Sie zum Beispiel *Geschl* als die erste Sortiervariable und *mind* als die zweite auswählen, wird innerhalb der Geschlechtskategorien nach Minderheit sortiert.
- Die Sortierreihenfolge beruht auf der durch das Gebietsschema definierten Reihenfolge (und stimmt nicht unbedingt mit der numerischen Reihenfolge der Zeichencodes überein). Standardmäßig wird das Gebietsschema des Betriebssystems verwendet. Sie können das Gebietsschema ändern, indem Sie im Menü “Bearbeiten” im Dialogfeld “Optionen” auf der Registerkarte “Allgemein” die Einstellung unter “Sprache” entsprechend anpassen.

Abbildung 9-1
Dialogfeld “Fälle sortieren”



So sortieren Sie Fälle:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
Fälle sortieren

▶ Wählen Sie mindestens eine Sortiervariable aus.

Variablen sortieren

Sie können die Variablen in der Datendatei anhand der Werte eines beliebigen Variablenattributs sortieren (z. B. nach Variablennamen, Datentyp, Messniveau), einschließlich benutzerdefinierter Variablenattribute.

- Die Werte können in aufsteigender oder absteigender Reihenfolge sortiert werden.
- Sie können die ursprüngliche Variablenreihenfolge (vor der Sortierung) in einem benutzerdefinierten Variablenattribut speichern.
- Für die Sortierung anhand der Werte benutzerdefinierter Variablenattribute können nur benutzerdefinierte Variablenattribute verwendet werden, die aktuell in der Variablenansicht sichtbar sind.

Weitere Informationen zu benutzerdefinierten Variablenattributen finden Sie unter Benutzerdefinierte Variablenattribute.

So sortieren Sie Variablen:

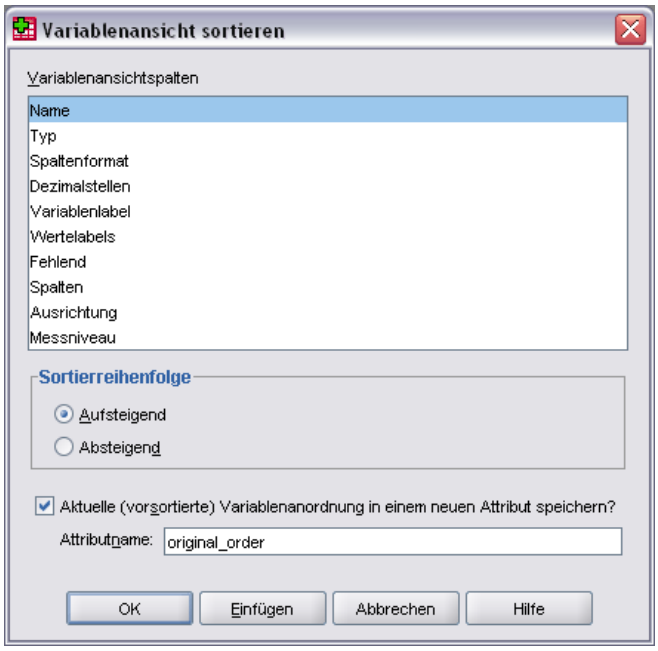
In der Variablenansicht im Daten-Editor:

- Klicken Sie mit der rechten Maustaste auf die Überschrift der Attributspalte und wählen Sie im Kontextmenü die Option Aufsteigend sortieren bzw. Absteigend sortieren.

oder

- Wählen Sie in einem beliebigen Fenster die folgenden Optionen aus den Menüs aus:
  Daten
  Variablen sortieren
- Wählen Sie das Attribut aus, das Sie zum Sortieren der Variablen verwenden möchten.
- Wählen Sie die Sortierreihenfolge (aufsteigend oder absteigend).

Abbildung 9-2
Dialogfeld "Variablenansicht sortieren"



- Die Liste der Variablenattribute stimmt mit den Namen der Attributspalten überein, die in der Variablenansicht des Daten-Editors angezeigt werden.
- Sie können die ursprüngliche Variablenreihenfolge (vor der Sortierung) in einem benutzerdefinierten Variablenattribut speichern. Für jede Variable ist der Wert des Attributs eine ganze Zahl, die die Position vor der Sortierung angibt; so kann durch Sortieren der Variablen anhand des Werts dieses benutzerdefinierten Attributs die ursprüngliche Reihenfolge der Variablen wiederhergestellt werden.

Transponieren

Mit “Transponieren” können Sie eine neue Datendatei anlegen, in welcher die Zeilen und Spalten aus der ursprünglichen Datendatei transponiert wurden, sodass die Fälle (also die Zeilen) zu Variablen und die Variablen (also die Spalten) zu Fällen werden. SPSS erzeugt automatisch neue Variablennamen und zeigt eine Liste der neuen Variablennamen an.

- SPSS erzeugt automatisch eine neue String-Variable, *case_lbl*, welche den ursprünglichen Variablennamen enthält.
- Falls die Arbeitsdatei eine ID- oder Namensvariable mit eindeutigen Werten enthält, können Sie diese im Feld “Namensvariable” eintragen. Die Werte dieser Variablen werden dann als Variablennamen in der transponierten Datendatei verwendet. Falls es sich um eine numerische Variable handelt, beginnen die Variablennamen mit dem Buchstaben *V*, gefolgt von einem numerischen Wert.
- Benutzerdefinierte fehlende Werte werden in der transponierten Datendatei in systemdefinierte fehlende Werte umgewandelt. Um diese Werte beizubehalten, müssen Sie die Definition fehlender Werte in der Variablenansicht des Daten-Editors ändern.

So transponieren Sie Variablen und Fälle:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Daten
Transponieren...

▶ Wählen Sie eine oder mehrere Variablen aus, die in Fälle umgewandelt werden sollen.

Zusammenfügen von Datendateien

Mit SPSS können Sie Daten aus zwei Dateien auf zwei verschiedene Arten zusammenfügen. Sie verfügen über folgende Möglichkeiten:

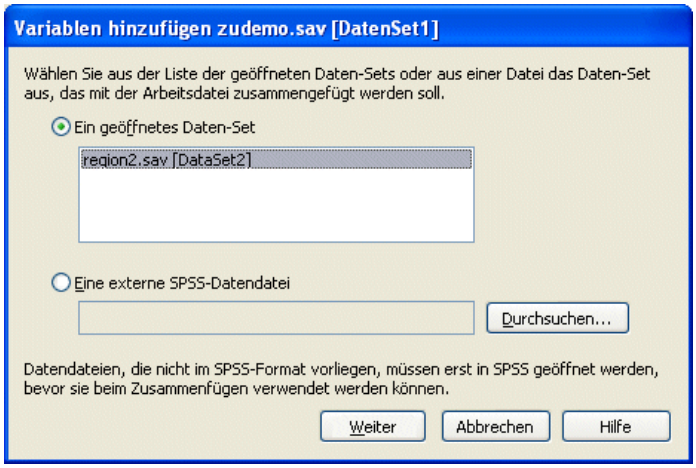
- Zusammenfügen der Arbeitsdatei mit einem weiteren geöffneten Daten-Set oder einer Datendatei im SPSS-Format, das bzw. die dieselben Variablen aber unterschiedliche Fälle enthält
- Zusammenfügen der Arbeitsdatei mit einem weiteren geöffneten Daten-Set oder einer Datendatei im SPSS-Format, das bzw. die dieselben Fälle aber unterschiedliche Variablen enthält

So fügen Sie Dateien zusammen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
Daten
Dateien zusammenfügen

▶ Wählen Sie Fälle hinzufügen oder Variablen hinzufügen aus.

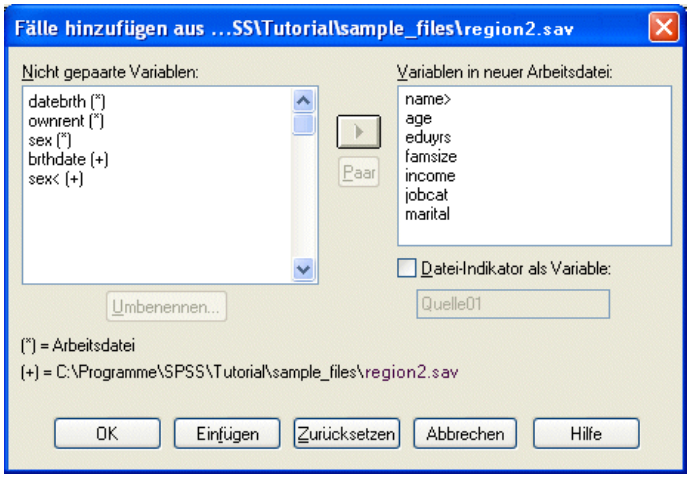
Abbildung 9-3
Auswahl der zusammenzufügenden Dateien



Fälle hinzufügen

Mithilfe von “Fälle hinzufügen” wird die Arbeitsdatei mit einem zweiten Daten-Set oder einer Datendatei im SPSS-Format zusammengefügt, das bzw. die dieselben Variablen (Spalten) aber unterschiedliche Fälle (Zeilen) enthält. So könnten Sie zum Beispiel dieselben Informationen für Kunden in zwei verschiedenen Verkaufsgebieten aufzeichnen und die Daten für jedes Gebiet in getrennten Dateien speichern. Das zweite Daten-Set kann eine externe Datendatei im SPSS-Format oder ein Daten-Set aus der aktuellen Sitzung sein.

Abbildung 9-4
Dialogfeld “Fälle hinzufügen”



Nicht gepaarte Variablen. Diese Variablen werden nicht in die neue, zusammengefügte Datendatei aufgenommen. Die Variablen aus der Arbeitsdatei sind mit einem Sternchen (*) gekennzeichnet. Die Variablen aus dem anderen Daten-Set sind mit einem Pluszeichen (+) gekennzeichnet. Diese Liste enthält in der Standardeinstellung folgende Einträge:

- Variablen, die nicht mit gleichem Namen in beiden Datendateien enthalten sind. Sie können Paare aus den nicht gepaarten Variablen bilden und diese in die neue, zusammengefügte Datei aufnehmen.
- Variablen, die als numerische Daten in der einen Datei und als String-Daten in der anderen Datei definiert sind. Numerische Variablen können nicht mit String-Variablen zusammengefügt werden.
- String-Variablen ungleicher Länge. Die definierte Länge einer String-Variablen muss in beiden Datendateien gleich sein.

Variablen in neuer Arbeitsdatei. Diese Variablen werden in die neue, zusammengefügte Datendatei aufgenommen. Die Standardeinstellung sieht vor, dass alle Variablen in die Liste aufgenommen werden, die sowohl in ihrem Namen als auch im Datentyp (numerisch oder String) übereinstimmen.

- Sie können Variablen, die nicht in die zusammengeführte Datei aufgenommen werden sollen, aus der Liste entfernen.
- Alle nicht gepaarten Variablen, die in die zusammengefügte Datei aufgenommen werden, enthalten fehlende Daten für die Fälle aus der Datei, die diese Variable nicht enthält.

Datei-Indikator als Variable. Gibt für jeden Fall die Quelldatei an. Diese Variable hat den Wert 0 für Fälle aus der Arbeitsdatei und den Wert 1 für Fälle aus der externen Datendatei.

So fügen Sie Datendateien mit denselben Variablen und unterschiedlichen Fällen zusammen:

▶ Öffnen Sie mindestens eine der Datendateien, die Sie zusammenfügen möchten. Wenn Sie mehrere Daten-Sets geöffnet haben, legen Sie eines der zusammenzufügenden Daten-Sets als Arbeitsdatei fest. Die Fälle aus dieser Datei werden zuerst in die neue, zusammengefügte Datendatei übernommen.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Dateien zusammenfügen
    Fälle hinzufügen

▶ Wählen Sie das Daten-Set oder die Datendatei im SPSS-Format aus, das bzw. die Sie mit der Arbeitsdatei zusammenfügen möchten.

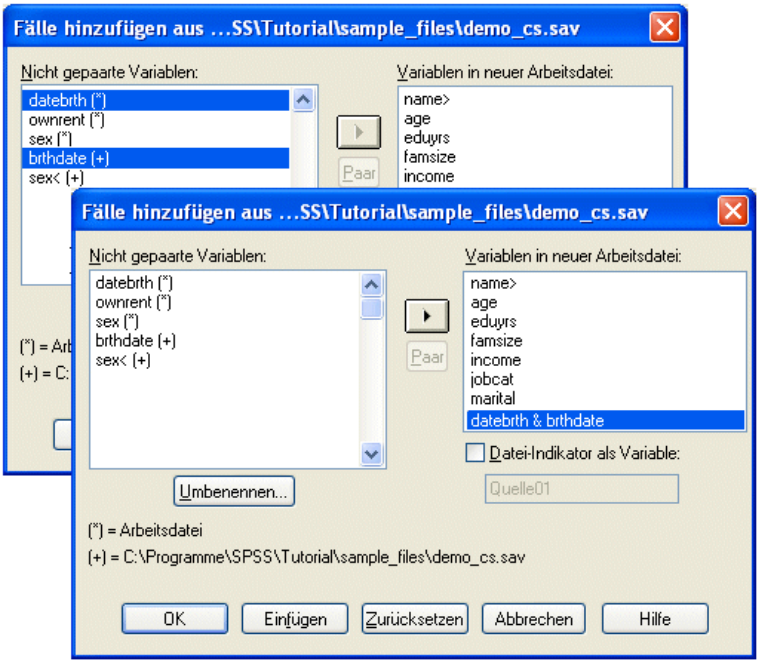
▶ Entfernen Sie alle Variablen, die nicht übernommen werden sollen, aus der Liste “Variablen in neuer Arbeitsdatei”.

▶ Fügen Sie gegebenenfalls die Variablenpaare aus der Liste “Nicht gepaarte Variablen” hinzu, die dieselben Informationen unter verschiedenen Variablennamen in den beiden Dateien enthalten. So könnte zum Beispiel das Geburtsdatum in der einen Datei den Variablennamen *Gebdat* haben und in der anderen Datei den Namen *DatGeb*.

So wählen Sie ein Paar nicht gepaarter Variablen aus:

- Klicken Sie in der Liste “Nicht gepaarte Variablen” auf eine Variable.
- Klicken Sie bei gedrückter STRG-Taste auf die andere Variable in der Liste. (Drücken Sie dazu gleichzeitig auf die STRG-Taste und die linke Maustaste.)
- Klicken Sie auf Paar, um das Variablenpaar in die Liste “Variablen in neuer Arbeitsdatei” zu verschieben. (Der Variablenname aus der Arbeitsdatei wird in der zusammengefügten Datei als Variablenname verwendet.)

Abbildung 9-5
Auswählen von Variablenpaaren durch Klicken mit der Maus bei gedrückter STRG-Taste



Fälle hinzufügen: Umbenennen

Sie können Variablen sowohl in der Arbeitsdatei als auch im anderen Daten-Set umbenennen, bevor Sie diese aus der Liste der nicht gepaarten Variablen in die Liste der Variablen verschieben, die in die zusammengeführte Datei aufgenommen werden. Mit dem Umbenennen von Variablen erreichen Sie folgendes:

- Sie können die Variablennamen aus dem anderen Daten-Set anstelle der Namen aus der Arbeitsdatei für Variablenpaare verwenden.
- Sie können zwei Variablen mit demselben Namen, aber unterschiedlichen Typen oder verschiedenen String-Längen aufnehmen. Wenn Sie zum Beispiel sowohl die numerische Variable *Geschl* aus der Arbeitsdatei als auch die String-Variable *Geschl* aus dem anderen Daten-Set aufnehmen möchten, müssen Sie zuerst eine der beiden Variablen umbenennen.

Fälle hinzufügen: Informationen aus dem Datenlexikon

Alle vorhandenen Informationen aus dem Datenlexikon der Arbeitsdatei (Variablen- und Wertelabels, benutzerdefinierte fehlende Werte und Anzeigenformate) werden der zusammengefügten Datendatei zugewiesen.

- Wenn für eine Variable keine Informationen im Datenlexikon der Arbeitsdatei definiert sind, werden die entsprechenden Informationen aus dem Datenlexikon des anderen Daten-Sets verwendet.
- Wenn die Arbeitsdatei definierte Wertelabels oder benutzerdefinierte fehlende Werte für eine Variable enthält, werden alle weiteren Wertelabels oder benutzerdefinierten fehlenden Werte für diese Variable im anderen Daten-Set ignoriert.

Zusammenfügen von mehr als zwei Datenquellen

Mithilfe der Befehlssyntax können Sie bis zu 50 Daten-Sets und/oder Datendateien zusammenfügen. Weitere Informationen finden Sie unter dem Befehl `ADD FILES` in der *Command Syntax Reference* (verfügbar über das Menü “Hilfe”).

Variablen hinzufügen

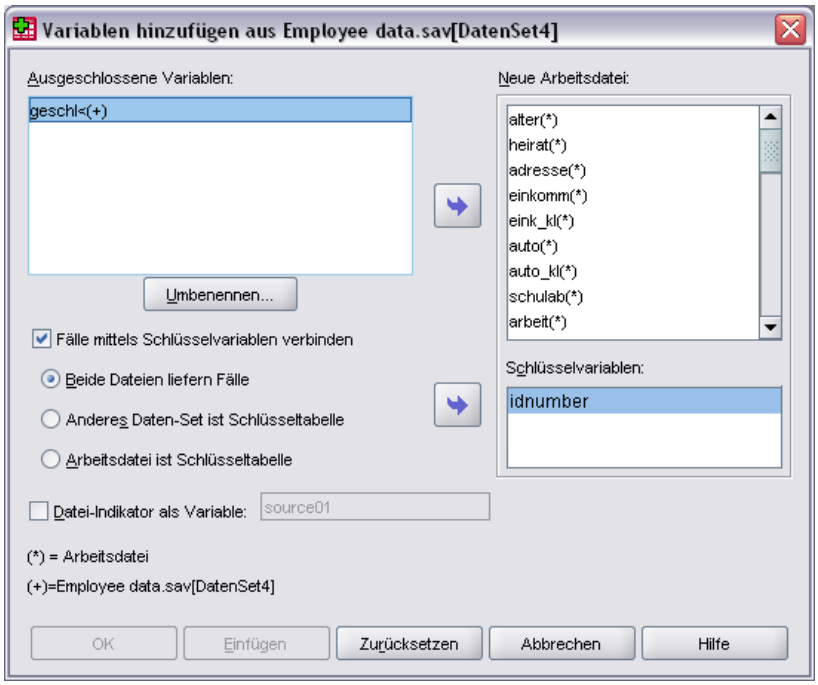
Mithilfe von “Variablen hinzufügen” wird die Arbeitsdatei mit einem weiteren geöffneten Daten-Set oder einer Datendatei im SPSS-Format zusammengefügt, das bzw. die dieselben Fälle (Zeilen) aber verschiedene Variablen (Spalten) enthält. So könnten Sie zum Beispiel eine Datendatei mit Ergebnissen vor einem Test mit einer Datei zusammenfügen, welche die Ergebnisse nach einem Test enthält.

- Die Fälle müssen in beiden Daten-Sets in derselben Reihenfolge sortiert sein.

- Wenn eine oder mehrere Schlüsselvariablen zum Zuordnen der Fälle verwendet werden, müssen die beiden Daten-Sets in aufsteigender Reihenfolge der Schlüsselvariable(n) sortiert sein.
- Variablennamen in der zweiten Datendatei, die mit Variablennamen in der Arbeitsdatei übereinstimmen, werden in der Standardeinstellung ausgeschlossen, da bei "Variablen hinzufügen" davon ausgegangen wird, dass diese Variablen dieselben Informationen enthalten.

Datei-Indikator als Variable. Gibt für jeden Fall die Quelldatei an. Diese Variable hat den Wert 0 für Fälle aus der Arbeitsdatei und den Wert 1 für Fälle aus der externen Datendatei.

Abbildung 9-6
Dialogfeld "Variablen hinzufügen"



Ausgeschlossene Variablen. Diese Variablen werden nicht in die neue, zusammengefügte Datendatei aufgenommen. Diese Liste enthält in der Standardeinstellung alle Variablennamen aus dem anderen Daten-Set, die mit Variablennamen aus der Arbeitsdatei übereinstimmen. Die Variablen aus der Arbeitsdatei sind mit einem Sternchen (*) gekennzeichnet. Die Variablen aus dem anderen Daten-Set sind mit einem Pluszeichen (+) gekennzeichnet. Wenn Sie eine ausgeschlossene Variable mit einem doppelt belegten Namen in die zusammengefügte Datei aufnehmen möchten, können Sie diese Variable umbenennen und sie der Liste einzuschließender Variablen hinzufügen.

Neue Arbeitsdatei. Diese Variablen werden in das neue, zusammengefügte Daten-Set aufgenommen. In der Standardeinstellung werden alle eindeutigen Variablennamen in beiden Daten-Sets in die Liste aufgenommen.

Schlüsselvariablen. Wenn einige Fälle in einem Daten-Set keine Entsprechung im anderen Daten-Set aufweisen (das heißt, einige Fälle fehlen in einem Daten-Set), verwenden Sie Schlüsselvariablen zum Identifizieren und richtigen Zuordnen der Fälle in den beiden Daten-Sets. Schlüsselvariablen können Sie außerdem bei Tabellenindexdateien verwenden.

- Die Schlüsselvariablen müssen in beiden Daten-Sets unter demselben Namen aufgeführt sein.
- Beide Daten-Sets müssen in aufsteigender Reihenfolge der Schlüsselvariablen sortiert sein, wobei die Variablen in der Liste "Schlüsselvariablen" entsprechend der Sortierfolge geordnet sein müssen.
- Fälle, die nicht in den Schlüsselvariablen übereinstimmen, werden in die zusammengefügte Datei aufgenommen, aber sie werden nicht mit Fällen aus der anderen Datei verknüpft. Fälle ohne Entsprechung enthalten nur Werte für die Variablen in der Datei, aus der sie stammen. Die Variablen aus der anderen Datei enthalten den systemdefinierten fehlenden Wert.

Anderes Daten-Set oder Arbeitsdatei ist Schlüsseltabelle. Eine Schlüsseltabelle oder **Tabellenindexdatei**, ist eine Datei, deren Daten für jeden Fall mehreren Fällen der anderen Datendatei zugeordnet werden können. Wenn eine Datei zum Beispiel Informationen über einzelne Familienmitglieder (beispielsweise Geschlecht, Alter und Bildungsstand) und die andere Datei übergreifende Informationen zur ganzen Familie enthält (beispielsweise das Gesamteinkommen, die Familiengröße und den Wohnort), können Sie die Datei mit den Familiendaten als Tabellenindexdatei verwenden. In der zusammengefügten Datendatei können Sie dann jedem einzelnen Familienmitglied die gemeinsamen Familiendaten zuweisen.

So fügen Sie Dateien mit gleichen Fällen, aber unterschiedlichen Variablen zusammen:

▶ Öffnen Sie mindestens eine der Datendateien, die Sie zusammenfügen möchten. Wenn Sie mehrere Daten-Sets geöffnet haben, legen Sie eines der zusammenzufügenden Daten-Sets als Arbeitsdatei fest.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Dateien zusammenfügen
    Variablen hinzufügen...

▶ Wählen Sie das Daten-Set oder die Datendatei im SPSS-Format aus, das bzw. die Sie mit der Arbeitsdatei zusammenfügen möchten.

So wählen Sie Schlüsselvariablen aus:

▶ Wählen Sie aus der Liste "Ausgeschlossene Variablen" die Variablen der externen Datei (mit "+" markiert) aus.

▶ Aktivieren Sie das Optionsfeld Fälle mittels Schlüsselvariablen verbinden.

▶ Fügen Sie die Variablen der Liste "Schlüsselvariablen" hinzu.

Die Schlüsselvariablen müssen sowohl in der Arbeitsdatei als auch im anderen Daten-Set vorhanden sein. Beide Daten-Sets müssen in aufsteigender Reihenfolge der Schlüsselvariablen sortiert sein, wobei die Variablen in der Liste "Schlüsselvariablen" entsprechend der Sortierfolge geordnet sein müssen.

Variablen hinzufügen: Umbenennen

Sie können Variablen sowohl in der Arbeitsdatei als auch im anderen Daten-Set umbenennen, bevor Sie diese in die Liste der Variablen verschieben, die in die zusammengeführte Datei aufgenommen werden. Dies empfiehlt sich, wenn Sie zwei Variablen mit gleichem Namen aufnehmen möchten, die verschiedene Informationen in den beiden Dateien enthalten.

Zusammenfügen von mehr als zwei Datenquellen

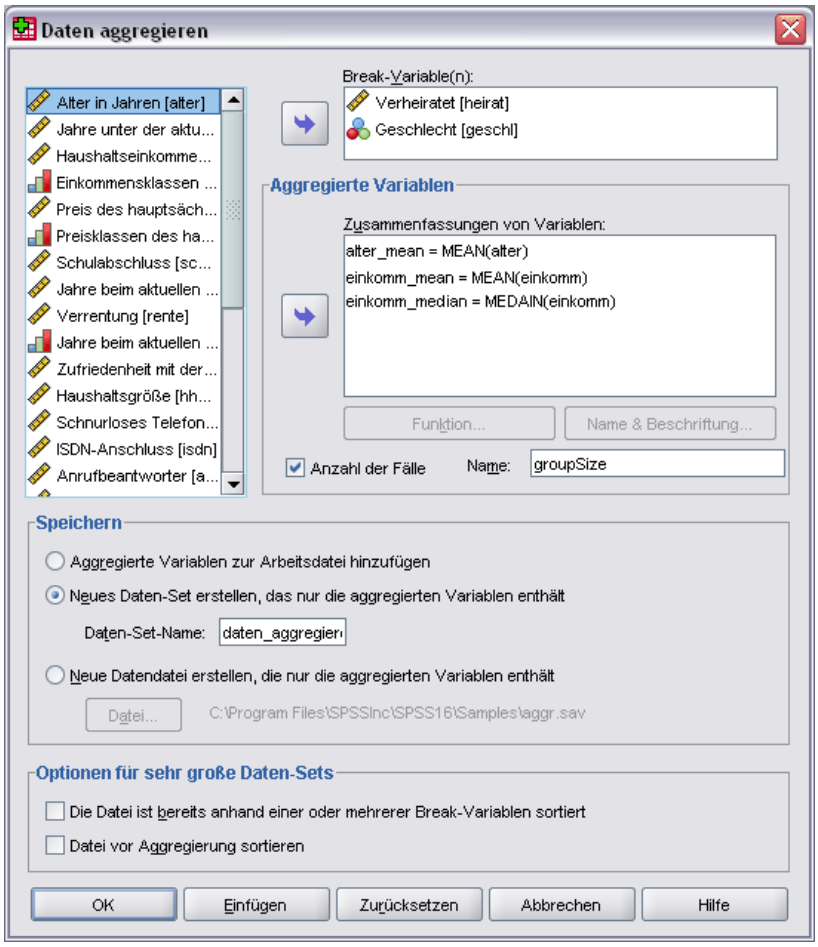
Mithilfe der Befehlssyntax können Sie bis zu 50 Daten-Sets und/oder Datendateien zusammenfügen. Weitere Informationen finden Sie unter dem Befehl `MATCH FILES` in der *Command Syntax Reference* (verfügbar über das Menü “Hilfe”).

Daten aggregieren

Mit “Daten aggregieren” werden Fallgruppen in der Arbeitsdatei zu einzelnen Fällen kombiniert; hierbei wird eine neue, aggregierte Datei angelegt oder es werden neue Variablen in der Arbeitsdatei angelegt, die aggregrierte Daten enthalten. Die Fälle werden nach einem oder mehreren Werten von Break-Variablen (Gruppenvariablen) aggregiert.

- Wenn Sie eine neue, aggregierte Datendatei anlegen, enthält diese neue Datei je einen Fall für jede Gruppe, die in den Break-Variablen definiert sind. Liegt beispielsweise eine Break-Variable mit zwei Werten vor, enthält die neue Datendatei nur zwei Fälle.
- Wenn Sie Aggregierungsvariablen in die Arbeitsdatei aufnehmen, wird die Datendatei selbst nicht aggregiert. Jeder Fall mit denselben Werten für die Break-Variable(n) erhält dieselben Werte für die neuen Aggregierungsvariablen. Wenn beispielsweise nur eine Break-Variable *geschl* vorliegt, erhalten alle männlichen Personen denselben Wert für eine neue Aggregierungsvariable, die das Durchschnittsalter erfasst.

Abbildung 9-7
Dialogfeld "Daten aggregieren"



Break-Variable(n). Die Fälle werden auf der Basis der Break-Variablen gruppiert. Jede eindeutige Kombination von Break-Variablenwerten definiert eine Gruppe. Wenn Sie eine neue, aggregierte Datendatei erstellen, werden alle Break-Variablen in der neuen Datei unter dem bisherigen Namen und mit den vorhandenen Informationen aus dem Datenlexikon gespeichert. Die Break-Variablen können numerische Variablen oder String-Variablen sein.

Aggregierte Variablen. Quellvariablen werden in Verbindung mit Aggregierungsfunktionen zum Erzeugen der neuen Aggregierungsvariablen herangezogen. Der Name der Aggregierungsvariablen wird von einem optionalen Variablenlabel, dem Namen der Aggregierungsfunktion und der Quellvariablen in Klammern gefolgt.

Sie können die vorgegebenen Namen für die Aggregierungsvariablen mit neuen Variablennamen überschreiben, aussagekräftige Variablenlabels verwenden und die Funktionen zum Berechnen der aggregierten Datenwerte ändern. Sie können auch eine Variable anlegen, welche die Anzahl der Fälle in jeder Break-Gruppe enthält.

So aggregieren Sie eine Datendatei:

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Daten
  Aggregieren...
- ▶ Wählen Sie eine oder mehrere Break-Variablen aus, die definieren, wie die Fälle zum Erzeugen der aggregierten Daten gruppiert werden.
- ▶ Wählen Sie mindestens eine Aggregierungsvariable aus.
- ▶ Wählen Sie für jede Aggregierungsvariable eine Aggregierungsfunktion aus.

Speichern von aggregierten Ergebnissen

Sie können wahlweise Aggregierungsvariablen in die Arbeitsdatei aufnehmen oder eine neue, aggregierte Datendatei erzeugen.

- **Aggregierte Variablen zur Arbeitsdatei hinzufügen.** Neue Variablen, die auf Aggregierungsfunktionen beruhen, werden zur Arbeitsdatei hinzugefügt. Die Datendatei selbst wird nicht aggregiert. Jeder Fall mit demselben Wert bzw. denselben Werten der Break-Variablen erhält die gleichen Werte für die neuen Aggregatvariablen.
- **Neues Daten-Set erstellen, das nur die aggregierten Variablen enthält.** Speichert die aggregierten Daten in ein neues Daten-Set in der aktuellen Sitzung. Das Daten-Set enthält die Break-Variablen, die die aggregierten Fälle bestimmen, und alle aggregierten Variablen, die durch Aggregierungsfunktionen definiert werden. Die Arbeitsdatei (das aktive Daten-Set) bleibt davon unberührt.
- **Neue Datendatei erstellen, die nur die aggregierten Variablen enthält.** Speichert die aggregierten Daten in einer externen Datendatei. Die Datei enthält die Break-Variablen, die die aggregierten Fälle bestimmen, und alle aggregierten Variablen, die durch Aggregierungsfunktionen definiert werden. Die Arbeitsdatei (das aktive Daten-Set) bleibt davon unberührt.

Sortieroptionen für umfangreiche Datendateien

Bei äußerst umfangreichen Datendateien empfiehlt es sich, die Dateien vor der Aggregierung zu sortieren.

Datei ist bereits nach Break-Variablen sortiert. Wenn die Daten bereits nach den Werten der Break-Variablen sortiert wurden, sorgt diese Option für einen schnelleren Ablauf der Prozedur und geringeren Speicherplatzbedarf. Diese Option sollte mit Bedacht verwendet werden.

- Die Daten müssen nach den Werten der Break-Variablen sortiert werden, und zwar in derselben Reihenfolge wie die Break-Variablen, die für die Funktion “Daten aggregieren” angegeben wurden.
- Wenn Sie Variablen in die Arbeitsdatei aufnehmen, wählen Sie diese Option nur dann aus, wenn die Daten anhand der Werte der Break-Variablen in aufsteigender Reihenfolge sortiert sind.

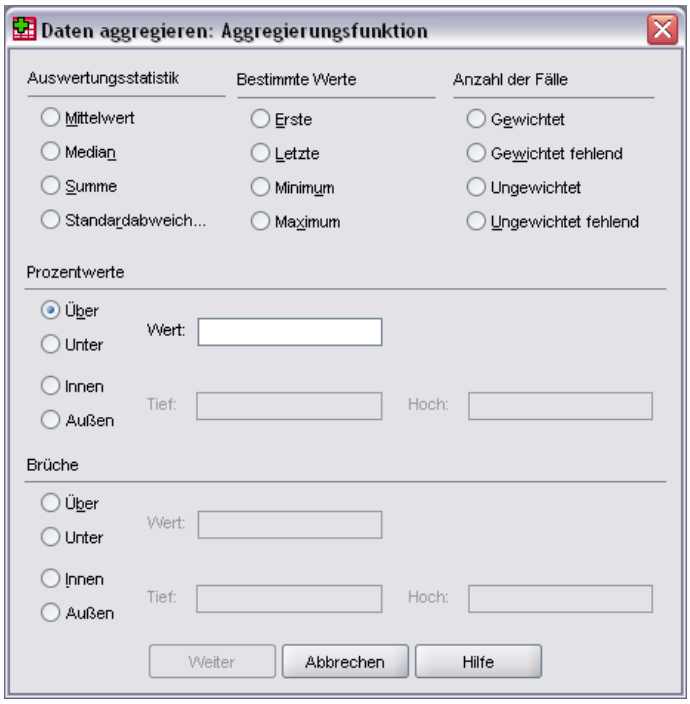
Datei vor Aggregierung sortieren. In sehr seltenen Fällen kann es bei großen Datendateien nötig sein, die Datendatei vor dem Aggregieren nach den Werten der Break-Variablen zu sortieren. Diese Option ist nur dann angeraten, wenn Speicher- bzw. Leistungsprobleme auftreten.

Daten aggregieren: Aggregierungsfunktion

In diesem Dialogfeld legen Sie die Funktion fest, die zum Berechnen der aggregierten Datenwerte für die Variablen verwendet wird, die in der Liste “Variablen aggregieren” aus dem Dialogfeld “Daten aggregieren” ausgewählt wurden. Folgende Aggregierungsfunktionen stehen zur Verfügung:

- Auswertungsfunktionen für numerische Variablen, einschließlich Mittelwert, Median, Standardabweichung und Summe.
- Anzahl der Fälle, einschließlich ungewichtet, gewichtet, nichtfehlend und fehlend.
- Prozentsatz oder Anteil von Werten über oder unter einem festgelegten Wert.
- Prozentsatz oder Anteil von Werten innerhalb oder außerhalb eines festgelegten Bereichs.

Abbildung 9-8
Dialogfeld “Daten aggregieren: Aggregierungsfunktion”



Daten aggregieren: Variablenname und -label

Mit “Daten aggregieren” werden den aggregierten Variablen in der neuen Datendatei vorgegebene Variablennamen zugewiesen. In diesem Dialogfeld können Sie den Variablennamen der ausgewählten Variablen in der Liste “Variablen aggregieren” ändern und ein aussagekräftiges Variablenlabel angeben. Für weitere Informationen siehe Variablennamen in Kapitel 5 auf S. 83.

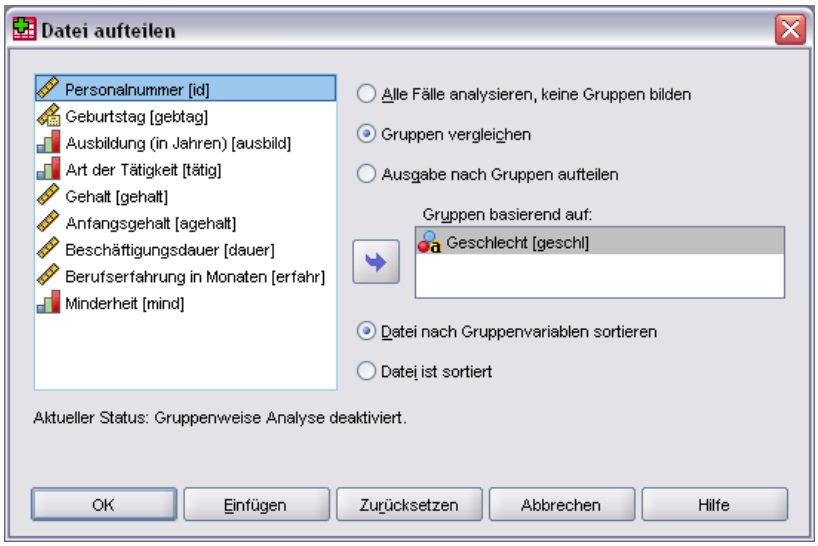
Abbildung 9-9
Dialogfeld "Daten aggregieren: Variablenname und -label"

Datei aufteilen

Mit "Datei aufteilen" wird die Datendatei in separate Gruppen für die Analyse aufgeteilt, und zwar nach dem Wert einer oder mehrerer Gruppenvariablen. Wenn Sie mehrere Gruppenvariablen auswählen, werden die Fälle nach jeder Variablen innerhalb von Kategorien der vorhergehenden Variablen in der Liste "Gruppen basierend auf" geordnet. Wenn Sie zum Beispiel *Geschl* als erste Sortiervariable auswählen und *Minder* als zweite, wird beim Sortieren innerhalb der Geschlechtskategorien nach Minderheit sortiert.

- Sie können bis zu acht Gruppenvariablen festlegen.
- Bei langen String-Variablen (String-Variablen mit mehr als acht Byte) gelten jeweils acht Byte als eine Variable, bis zur Obergrenze von acht Gruppenvariablen.
- Die Fälle müssen nach den Werten der Gruppenvariablen und in derselben Reihenfolge sortiert werden wie die in der Liste "Gruppen basierend auf" aufgeführten Variablen. Wenn die Datendatei noch nicht sortiert ist, klicken Sie auf Datei nach Gruppenvariablen sortieren.

Abbildung 9-10
Dialogfeld "Datei aufteilen"



Gruppen vergleichen. Die Gruppen der aufgeteilten Datei werden zu Vergleichszwecken zusammen angezeigt. Bei Pivot-Tabellen wird eine einzelne Pivot-Tabelle angelegt. Die Variablen zum Aufteilen der Datei können zwischen den Tabellendimensionen verschoben werden. Bei

Diagrammen wird zu jeder Gruppe der aufgeteilten Datei ein separates Diagramm erstellt. Die Diagramme werden zusammen im Viewer angezeigt.

Ausgabe nach Gruppen aufteilen. Die Ergebnisse jeder Prozedur werden für jede Gruppe einer aufgeteilten Datei separat angezeigt.

So teilen Sie eine Datendatei für die Analyse auf:

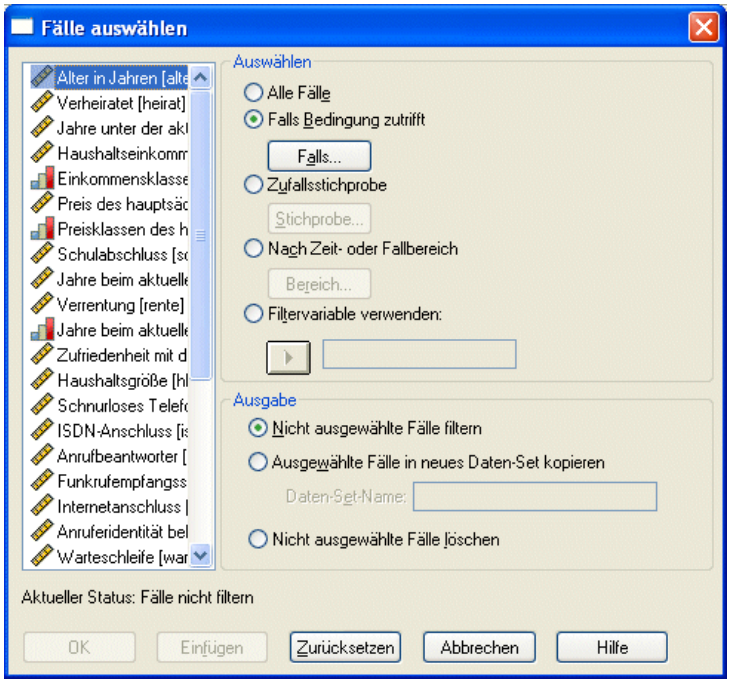
- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Daten
  Datei aufteilen...
- Aktivieren Sie das Optionsfeld Gruppen vergleichen oder Ausgabe nach Gruppen aufteilen.
- Wählen Sie eine oder mehrere Gruppenvariablen aus.

Fälle auswählen

Im Dialogfeld “Fälle auswählen” stehen Ihnen mehrere Methoden zum Auswählen einer Untergruppe von Fällen zur Verfügung. Diese Methoden basieren auf Kriterien, die unter anderem Variablen und komplexe Ausdrücke zulassen. Sie können auch eine Zufallsstichprobe aus den Fällen auswählen. Die Kriterien zum Festlegen der Untergruppen können folgende Elemente enthalten:

- Variablenwerte und -bereiche
- Datums- und Zeitbereiche
- Fallnummern (Zeilennummern)
- Arithmetische Ausdrücke
- Logische Ausdrücke
- Funktionen

Abbildung 9-11
Dialogfeld "Fälle auswählen"



Alle Fälle. Deaktiviert die Filterung der Fälle und verwendet alle Fälle.

Falls Bedingung zutrifft. Verwendet eine Bedingung zur Auswahl von Fällen. Wenn das Ergebnis der Bedingung wahr ist, wird der Fall ausgewählt. Wenn das Ergebnis falsch oder fehlend ist, wird der Fall nicht ausgewählt.

Zufallsstichprobe. Auswahl einer Zufallsstichprobe, deren Fallzahl näherungsweise durch einen Prozentsatz oder durch die genaue Anzahl vorgegeben werden kann.

Nach Zeit- oder Fallbereich. Wählt Fälle auf der Grundlage eines Bereichs von Fallnummern oder eines Datums- bzw. Zeitbereichs aus.

Filtervariable verwenden. Verwendet die ausgewählte, numerische Variable aus der Datendatei als Filtervariable. Fälle mit einem anderen Wert als 0 oder fehlend für die Filtervariable werden ausgewählt.

Ausgabe

In diesem Abschnitt wird die Behandlung von nicht ausgewählten Fällen festgelegt. Die folgenden Optionen stehen für die Behandlung nicht ausgewählter Fälle zur Auswahl:

- **Nicht ausgewählte Fälle filtern.** Nicht ausgewählte Fälle werden nicht in die Analyse aufgenommen, verbleiben jedoch im Daten-Set. Sie können die nicht ausgewählten Fälle später in der Sitzung verwenden, wenn Sie die Filterfunktion deaktivieren. Wenn Sie eine Zufallsstichprobe oder Fälle anhand eines bedingten Ausdrucks auswählen, wird die Variable *filter_$* mit dem Wert 1 für ausgewählte Fälle und dem Wert 0 für nicht ausgewählte Fälle erzeugt.

- **Ausgewählte Fälle in neues Daten-Set kopieren.** Die ausgewählten Fälle werden in ein neues Daten-Set kopiert, das ursprüngliche Daten-Set bleibt unverändert. Nicht ausgewählte Fälle werden nicht in das neue Daten-Set aufgenommen. Sie verbleiben im ursprünglichen Zustand im ursprünglichen Daten-Set.
- **Nicht ausgewählte Fälle löschen.** Nicht ausgewählte Fälle werden aus dem Daten-Set gelöscht. Gelöschte Fälle können nur wiederhergestellt werden, indem Sie die Datei ohne Speichern der Änderungen schließen und sie dann erneut öffnen. Wenn Sie die Änderungen in der Datendatei speichern, werden die Fälle dauerhaft gelöscht.

Anmerkung: Wenn Sie nicht ausgewählte Fälle löschen und die Datei speichern, können die Fälle nicht wiederhergestellt werden.

So wählen Sie eine Teilmenge von Fällen aus:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Fälle auswählen...

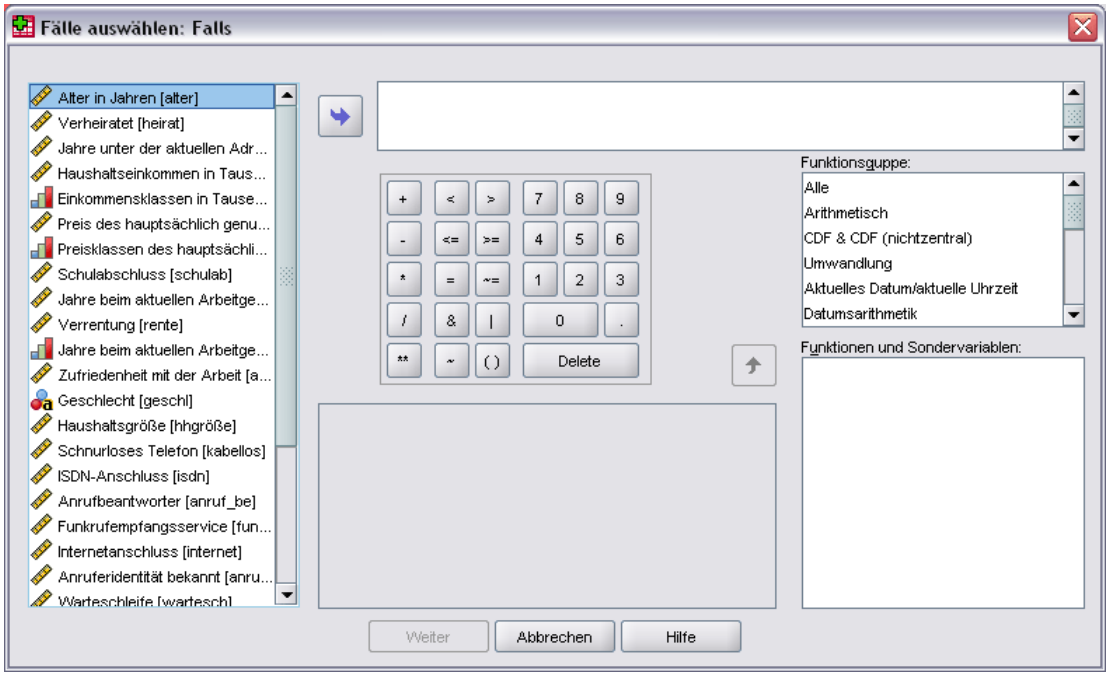
▶ Wählen Sie eine der Methoden zum Auswählen von Fällen.

▶ Geben Sie die Kriterien für die Auswahl der Fälle an.

Fälle auswählen: Falls

In diesem Dialogfeld können Sie anhand eines bedingten Ausdrucks eine Teilmenge der Fälle auswählen. Ein bedingter Ausdruck gibt für jeden Fall den Wert *Wahr*, *Falsch* oder *Fehlend* zurück.

Abbildung 9-12
Dialogfeld "Fälle auswählen: Falls"

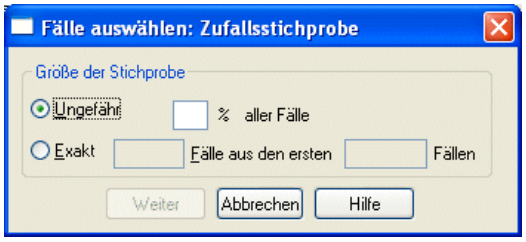


- Wenn das Ergebnis eines bedingten Ausdrucks *Wahr* ist, wird der Fall in die ausgewählte Untergruppe aufgenommen.
- Wenn das Ergebnis eines bedingten Ausdrucks *Falsch* oder *Fehlend* lautet, wird der Fall nicht in die ausgewählte Untergruppe aufgenommen.
- In den meisten bedingten Ausdrücke wird mindestens einer der sechs Vergleichsoperatoren (<, >, <=, >=, = und ~=) verwendet. Diese sind auf der Rechentastatur verfügbar.
- Bedingte Ausdrücke können Variablennamen, Konstanten, arithmetische Operatoren, numerische und andere Funktionen, logische Variablen und Vergleichsoperatoren enthalten.

Fälle auswählen: Zufallsstichprobe

In diesem Dialogfeld können Sie eine Zufallsstichprobe nach einem ungefähren Prozentsatz oder einer genauen Anzahl von Fällen auswählen. Für die Stichproben erfolgt keine Ersetzung, sodass ein Fall jeweils nur einmal ausgewählt werden kann.

Abbildung 9-13
Dialogfeld "Fälle auswählen: Zufallsstichprobe"



Ungefähr. Erstellt eine Zufallsstichprobe, die ungefähr den angegebenen Prozentsatz aller Fälle enthält. Da SPSS für jeden Fall eine unabhängige Pseudo-Zufallsentscheidung trifft, entspricht der Prozentsatz der tatsächlich ausgewählten Fälle nur ungefähr dem angegebenen Prozentwert. Je mehr Fälle sich in der Datendatei befinden, desto eher entspricht der Prozentsatz ausgewählter Fälle dem angegebenen Prozentsatz.

Exakt. Geben Sie die gewünschte Anzahl der Fälle ein. Sie müssen außerdem die Anzahl der Fälle angeben, aus denen die Stichprobe gezogen werden soll. Diese zweite Zahl muss kleiner oder gleich der Gesamtanzahl der Fälle in der Datendatei sein. Wenn die angegebene Anzahl die Gesamtanzahl der Fälle in der Datendatei übersteigt, enthält die Stichprobe entsprechend weniger Fälle als die geforderte Anzahl.

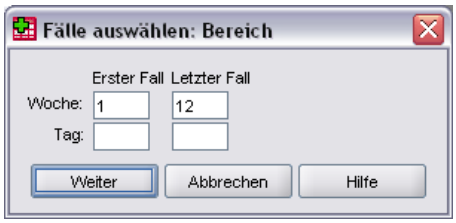
Fälle auswählen: Bereich

In diesem Dialogfeld können Sie Fälle anhand eines Bereichs von Fallnummern oder eines Zeitbereichs auswählen.

- Die Fallbereiche basieren auf der Zeilennummer, die im Daten-Editor angezeigt wird.
- Datums- und Zeitbereiche sind nur für **Zeitreihendaten** mit definierten Datumsvariablen verfügbar (Menü “Daten”, Befehl “Datum definieren”).

Abbildung 9-14
Dialogfeld “Fälle auswählen: Bereich” für Fallbereich (keine definierten Datumsvariablen)

Abbildung 9-15
Dialogfeld “Fälle auswählen: Bereich” für Zeitreihendaten mit definierten Datumsvariablen



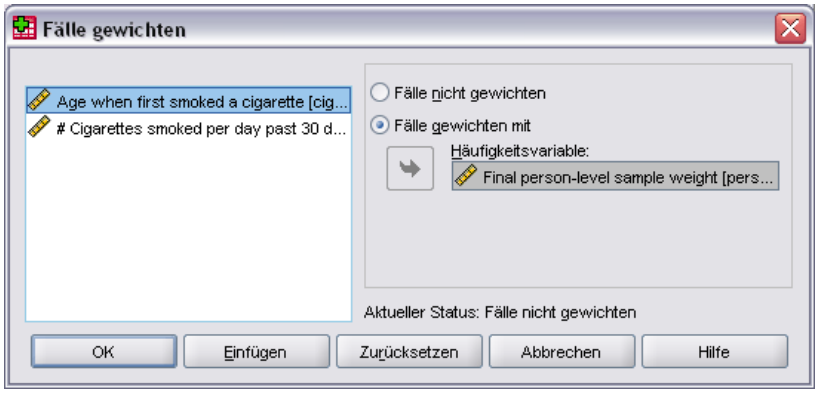
Fälle gewichten

Mit “Fälle gewichten” werden Fälle für die statistische Analyse unterschiedlich gewichtet (durch simulierte Replikation).

- Die Werte der Gewichtungsvariablen müssen der Anzahl der Beobachtungen entsprechen, die durch einzelne Fälle in der Datendatei dargestellt wird.

- Fälle, bei denen die Gewichtungsvariablen den Wert Null, einen negativen Wert oder einen fehlenden Wert aufweisen, werden von der Analyse ausgeschlossen.
- Gebrochene Zahlen sind zulässig. Sie werden nur verwendet, wenn dies sinnvoll ist, und höchstwahrscheinlich bei Fällen in Tabellen.

Abbildung 9-16
Dialogfeld "Fälle gewichten"



Sobald Sie eine Gewichtungsvariable zugewiesen haben, bleibt diese solange wirksam, bis Sie eine andere Gewichtungsvariable auswählen oder die Gewichtung ausschalten. Wenn Sie eine gewichtete Datendatei speichern, werden die Informationen für die Gewichtung zusammen mit der Datendatei gespeichert. Sie können die Gewichtung jederzeit aufheben, selbst wenn die Datei bereits in gewichteter Form gespeichert wurde.

Gewichtungen in der Prozedur "Kreuztabellen". In der Kreuztabellen-Prozedur gibt es mehrere Optionen für die Behandlung von Fallgewichtungen. Für weitere Informationen siehe Kreuztabellen: Zellen anzeigen in Kapitel 16 auf S. 295.

Gewichtungen bei Streudiagrammen und Histogrammen. Bei Streudiagrammen und Histogrammen gibt es eine Option zum Ein- und Ausschalten von Fallgewichtungen. Dies betrifft jedoch nicht die Fälle, bei denen die Gewichtungsvariable den Wert Null oder einen negativen oder fehlenden Wert aufweist. Diese Fälle bleiben selbst dann vom Diagramm ausgeschlossen, wenn Sie die Gewichtung aus dem Diagramm heraus ausschalten.

So gewichten Sie Fälle:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
Fälle gewichten...

▶ Aktivieren Sie das Optionsfeld Fälle gewichten mit.

▶ Wählen Sie eine Häufigkeitsvariable aus.

Die Werte der Häufigkeitsvariablen werden als Fallgewichtungen verwendet. So steht zum Beispiel ein Fall mit dem Wert 3 bei der Häufigkeitsvariablen in der gewichteten Datendatei für drei Fälle.

Umstrukturieren von Daten

Verwenden Sie den Assistenten für die Datenumstrukturierung, um Daten so umzustrukturieren, dass sie in der gewünschten Prozedur verwendet werden können. Der Assistent ersetzt die aktuelle Datei durch eine neue, umstrukturierte Datei. Der Assistent stellt Ihnen folgende Optionen zur Verfügung:

- Umstrukturieren ausgewählter Variablen in Fälle
- Umstrukturieren ausgewählter Fälle in Variablen
- Transponieren sämtlicher Daten

So strukturieren Sie Daten um:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Daten
  Umstrukturieren...

▶ Wählen Sie den Typ der Umstrukturierung aus, den Sie durchführen möchten.

▶ Wählen Sie die Daten aus, die umstrukturiert werden sollen.

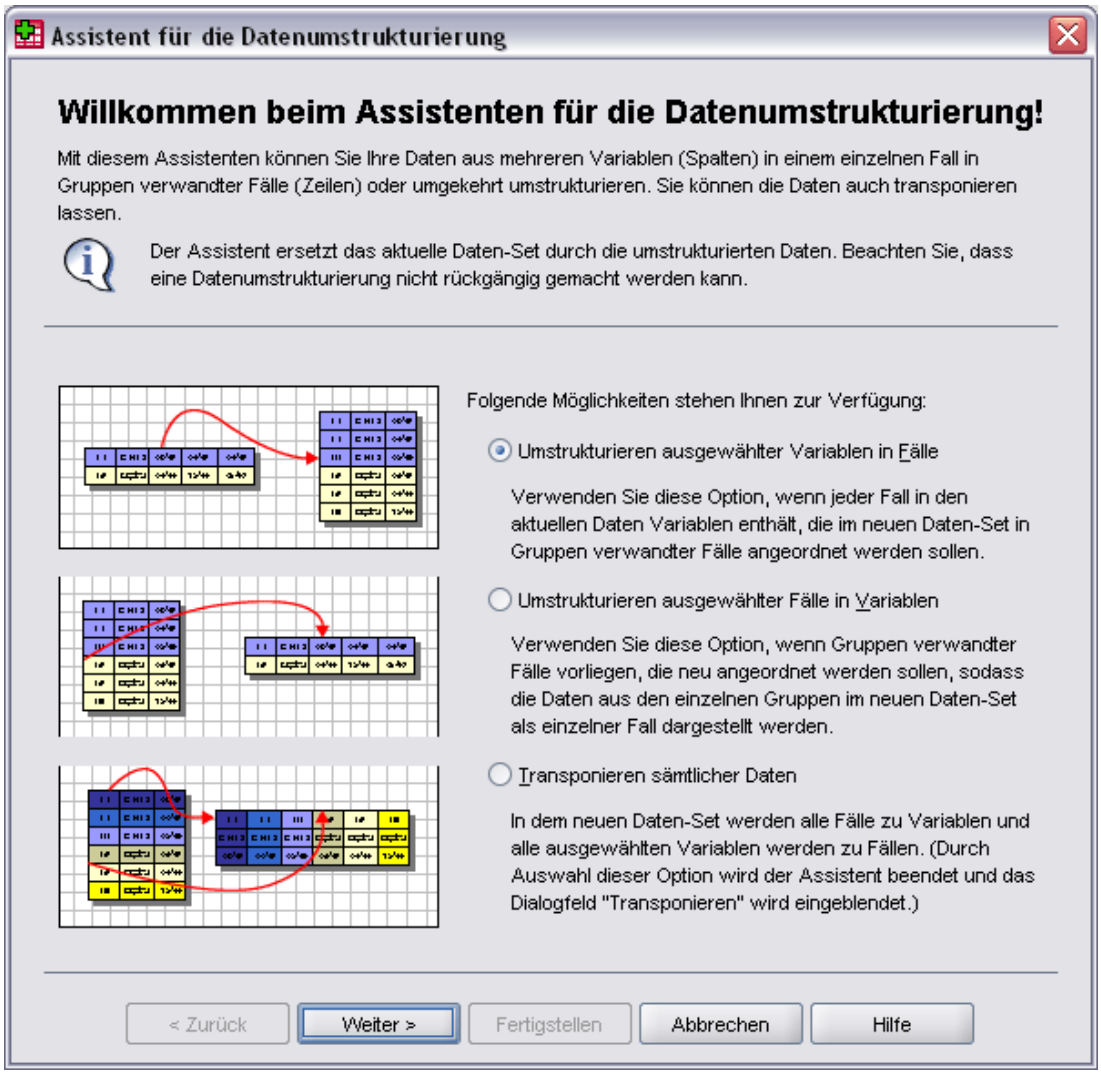
Die folgenden Optionen sind verfügbar:

- Erstellen von Bezeichnervariablen, mit deren Hilfe Sie einen Wert aus der neuen Datei zu einem Wert in der ursprünglichen Datei zurückverfolgen können
- Sortieren der Daten vor der Umstrukturierung
- Festlegen von Optionen für die neue Datei
- Einfügen der Befehlssyntax in ein Syntaxfenster

Assistent für die Datenumstrukturierung: Auswählen des Typs

Mit dem Assistenten für die Datenumstrukturierung können Sie Daten umstrukturieren. Wählen Sie im ersten Dialogfeld den gewünschten Umstrukturierungstyp aus.

Abbildung 9-17
Assistent für die Datenumstrukturierung:



- **Umstrukturieren ausgewählter Variablen in Fälle.** Wählen Sie diese Option aus, wenn die Daten Gruppen verwandter Spalten enthalten, die in der neuen Datendatei als Gruppen von Zeilen dargestellt werden sollen. Bei Auswahl dieser Option zeigt der Assistent die Schritte für **Variablen zu Fällen** an.
- **Umstrukturieren ausgewählter Fälle in Variablen.** Wählen Sie diese Option aus, wenn die Daten Gruppen verwandter Zeilen enthalten, die in der neuen Datendatei als Gruppen von Spalten dargestellt werden sollen. Bei Auswahl dieser Option zeigt der Assistent die Schritte für **Fälle zu Variablen** an.
- **Transponieren sämtlicher Daten.** Wählen Sie diese Option, wenn Sie die Daten vertauschen möchten. Dadurch werden alle Zeilen in den neuen Daten zu Spalten und alle Spalten zu Zeilen. Bei Auswahl dieser Option wird der Assistent für die Datenumstrukturierung geschlossen und das Dialogfeld “Daten transponieren” geöffnet.

Festlegen der Art der Umstrukturierung der Daten

Eine **Variable** enthält zu analysierende Informationen, z. B. eine Messung oder einen Wert. Ein **Fall** ist eine Beobachtung, z. B. eine Person. In einer *einfachen* Datenstruktur ist jede Variable eine einzelne Spalte in den Daten, und jeder Fall bildet eine einzelne Zeile. Wenn Sie beispielsweise Testergebnisse für alle Schüler einer Klasse messen, werden alle Messwerte in einer einzigen Spalte angezeigt, und jedem Schüler wird eine Zeile zugeordnet.

Beim Analysieren von Daten wird häufig untersucht, wie eine Variable in Abhängigkeit von einer bestimmten Bedingung variiert. Bei der Bedingung kann es sich um eine bestimmte zu erprobende Behandlung, eine demografische Gruppe, einen Zeitpunkt usw. handeln. In der Datenanalyse werden relevante Bedingungen häufig als **Faktoren** bezeichnet. Wenn Sie Faktoren analysieren, liegt eine *komplexe* Datenstruktur vor. Die Informationen zu einer Variablen können dabei in mehreren Datenspalten vorliegen (beispielsweise eine Spalte für jede Faktorstufe) oder es können in mehreren Zeilen Informationen zu einem Fall vorkommen (beispielsweise eine Zeile für jede Faktorstufe). Der Assistent für die Datenumstrukturierung unterstützt Sie bei der Umstrukturierung von Dateien mit einer komplexen Datenstruktur.

Welche Optionen Sie in diesem Assistenten auswählen, hängt von der Struktur der aktuellen Datei und der gewünschten Struktur der neuen Datei ab.

Wie sind die Daten in der aktuellen Datei angeordnet? Die aktuellen Daten können so angeordnet sein, dass Faktoren in einer *separaten* Variable (in Fallgruppen) oder *mit* der Variablen (in Variablengruppen) aufgezeichnet sind.

- **Fallgruppen.** Sind die Variablen und Bedingungen in der aktuellen Datei in verschiedenen Spalten aufgezeichnet? Beispiel:

Variable	Faktor
8	1
9	1
3	2
1	2

In diesem Beispiel handelt es sich bei den ersten zwei Zeilen um eine **Fallgruppe**, da sie miteinander verbunden sind. Sie enthalten Daten für dieselbe Faktorstufe. Wenn die Daten in dieser Weise strukturiert sind, wird der Faktor in der Datenanalyse von SPSS oft als **Gruppenvariable** bezeichnet.

- **Spaltengruppen.** Sind die Variablen und Bedingungen der aktuellen Datei in derselben Spalte aufgezeichnet? Beispiel:

var_1	var_2
8	3
9	1

In diesem Beispiel bilden die beiden Spalten eine **Variablengruppe** , da sie miteinander verbunden sind. Sie enthalten Daten für dieselbe Variable: *var_1* für Faktorstufe 1 und *var_2* für Faktorstufe 2. Wenn die Daten in dieser Weise strukturiert sind, wird der Faktor in der Datenanalyse von SPSS oft als **Messwiederholung** bezeichnet.

Wie sollen die Daten in der neuen Datei angeordnet werden? Dies hängt in der Regel von der Prozedur ab, mit der Sie die Daten analysieren möchten.

- **Prozeduren, die Fallgruppen erfordern.** Zum Durchführen von Analysen, die eine Gruppenvariable erfordern, müssen die Daten in Fallgruppen angeordnet sein. Beispiele hierfür sind *univariat*, *multivariat* und *Varianzkomponenten* mit dem allgemeinen linearen Modell, gemischten Modellen und OLAP-Würfeln sowie *unabhängige Stichproben* mit dem T-Test oder mit nichtparametrischen Tests. Wenn Sie diese Analysen durchführen möchten und Ihre Daten in Form von Variablengruppen strukturiert sind, wählen Sie Umstrukturieren ausgewählter Variablen in Fälle aus.
- **Prozeduren, die Variablengruppen erfordern.** Zum Analysieren von Messwiederholungen müssen die Daten in Variablengruppen angeordnet sein. Beispiele hierfür sind *Messwiederholungen* mit dem allgemeinen linearen Modell, die Analyse von *zeitabhängigen Kovariaten* mit der Cox-Regressionsanalyse, *gepaarte Stichproben* mit dem T-Test oder *verbundene Stichproben* mit nichtparametrischen Tests. Wenn Sie diese Analysen durchführen möchten und Ihre Daten in Form von Fallgruppen strukturiert sind, wählen Sie Umstrukturieren ausgewählter Fälle in Variablen aus.

Beispiel für die Umstrukturierung von Variablen zu Fällen

In diesem Beispiel sind die Testergebnisse für jeden Faktor (*A* und *B*) in verschiedenen Spalten aufgezeichnet.

Abbildung 9-18
Aktuelle Daten für "Variablen zu Fälle"

	score_a	score_b
1	1014.00	864.00
2	684.00	636.00
3	810.00	638.00

Es soll ein *T*-Test bei unabhängigen Stichproben durchgeführt werden. Sie verfügen über eine Spaltengruppe, die aus *score_a* und *score_b* besteht, nicht jedoch über die für die Prozedur erforderliche **Gruppenvariable**. Wählen Sie im Assistenten für die Datenumstrukturierung Umstrukturieren ausgewählter Variablen in Fälle aus, strukturieren Sie eine Variablengruppe in eine neue Variable mit der Bezeichnung *score* um, und erstellen Sie einen Index unter dem Namen *group*. Die neue Datendatei wird in der folgenden Abbildung dargestellt.

Abbildung 9-19
Neue umstrukturierte Daten für "Variablen zu Fälle"

	group	score
1	SCORE_A	1014.00
2	SCORE_B	864.00
3	SCORE_A	684.00
4	SCORE_B	636.00
5	SCORE_A	810.00
6	SCORE_B	638.00

Wenn Sie den *T*-Test bei unabhängigen Stichproben ausführen, können Sie nun *group* als Gruppenvariable verwenden.

Beispiel für die Umstrukturierung von Fällen zu Variablen

In diesem Beispiel werden die Testergebnisse für jedes Subjekt zweimal aufgezeichnet, und zwar vor und nach einer Behandlung.

Abbildung 9-20
Aktuelle Daten für "Fälle zu Variablen"

	id	wert	zeit
1	1	1014.00	bef
2	1	864.00	aft
3	2	684.00	bef
4	2	636.00	aft

Es soll ein *T*-Test bei gepaarten Stichproben durchgeführt werden. Die Daten sind in Fallgruppen angeordnet, es fehlen jedoch die **Messwiederholungen** für die Variablenpaare, die für die Prozedur erforderlich sind. Wählen Sie im Assistenten für die Datenumstrukturierung Umstrukturieren ausgewählter Fälle in Variablen aus, verwenden Sie *id* zum Identifizieren der Zeilengruppen in den aktuellen Daten und *zeit*, um die Variablengruppe in der neuen Datei zu erstellen.

Abbildung 9-21
Neue umstrukturierte Daten für "Fälle zu Variablen"

	id	wertnach	wertvorh
1	1	864.00	1014.00
2	2	636.00	684.00

Wenn Sie den *T*-Test bei gepaarten Stichproben ausführen, können Sie nun *Vorh* und *Nach* als Variablenpaar verwenden.

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Anzahl von Variablengruppen

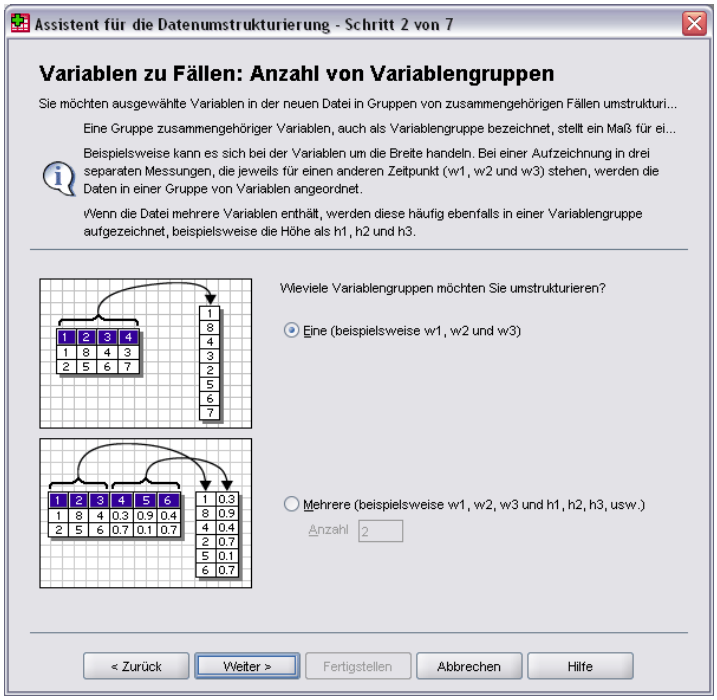
Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Variablengruppen in Zeilen umstrukturiert werden sollen.

Legen Sie in diesem Schritt fest, wie viele Variablengruppen der aktuellen Datei in der neuen Datei umstrukturiert werden sollen.

Wie viele Variablengruppen gibt es in der aktuellen Datei? Stellen Sie fest, wie viele Variablengruppen in den aktuellen Daten vorliegen. Eine Gruppe verbundener Spalten, auch **Variablengruppe** genannt, zeichnet verschiedene Messungen derselben Variablen in mehreren Spalten auf. Wenn es in den aktuellen Daten beispielsweise drei Spalten mit den Bezeichnungen *w1*, *w2* und *w3* zum Aufzeichnen der **Breite** gibt, stellen diese Spalten eine Variablengruppe dar. Wenn drei weitere Spalten *h1*, *h2* und *h3* zum Aufzeichnen der **Höhe** vorhanden sind, verfügen Sie über zwei Variablengruppen.

Wie viele Variablengruppen soll die neue Datei enthalten? Legen Sie fest, wie viele Variablengruppen in der neuen Datendatei dargestellt werden sollen. Sie müssen nicht notwendigerweise alle Variablengruppen in die neue Datei umstrukturieren.

Abbildung 9-22
Assistent für die Datenumstrukturierung: Anzahl der Variablengruppen, Schritt 2



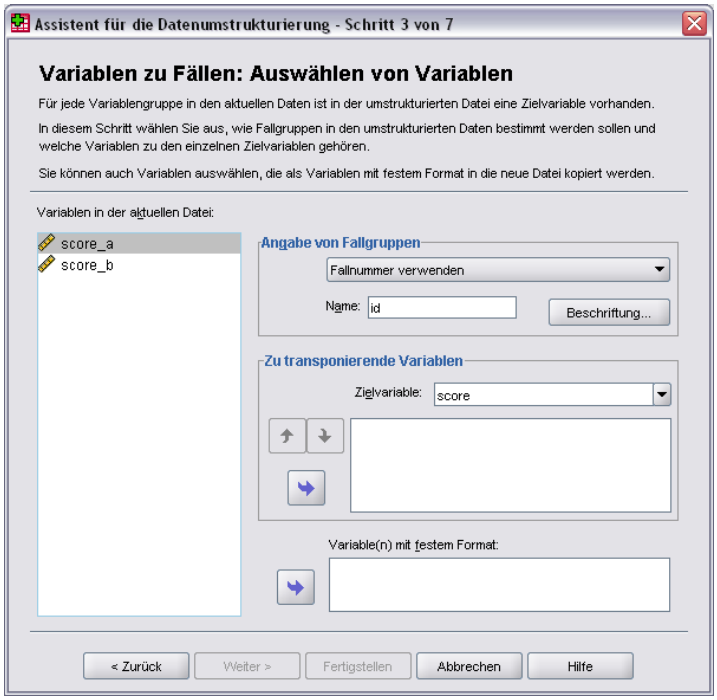
- **Eine.** Der Assistent erstellt in der neuen Datei eine einzige umstrukturierte Variable aus einer Variablengruppe der aktuellen Datei.
- **Mehrere.** Der Assistent erstellt in der neuen Datei mehrere umstrukturierte Variablen. Die angegebene Anzahl wirkt sich auf den nächsten Schritt aus, in dem der Assistent automatisch die angegebene Anzahl von neuen Variablen erstellt.

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Auswählen Variablen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Variablengruppen in Zeilen umstrukturiert werden sollen.

In diesem Schritt geben Sie an, wie die Variablen der aktuellen Datei in der neuen Datei verwendet werden sollen. Sie können auch eine Variable erstellen, welche die Zeilen in der neuen Datei identifiziert.

Abbildung 9-23
Assistent für die Datenumstrukturierung: Variablen auswählen, Schritt 3



Wie sollen die neuen Zeilen angegeben werden? Sie können in der neuen Datei eine Variable erstellen, welche die Zeile in der aktuellen Datei identifiziert, die zum Erstellen einer Gruppe von neuen Zeilen verwendet wurde. Diese Bezeichnervariable kann aus fortlaufenden Fallnummern oder den Werten der Variablen bestehen. Verwenden Sie die Steuerelemente in "Angabe von Fallgruppen", um die Bezeichnervariable in der neuen Datei zu definieren. Klicken Sie auf eine Zelle, um den vorgegebenen Variablennamen zu ändern und ein beschreibendes Variablenlabel für die Bezeichnervariable einzugeben.

Was soll in der neuen Datei umstrukturiert werden? Im vorhergehenden Schritt haben Sie angegeben, wie viele Variablengruppen umstrukturiert werden sollen. Der Assistent erstellt für jede Gruppe eine neue Variable. Die Werte für die Variablengruppe werden in dieser Variablen in der neuen Datei dargestellt. Verwenden Sie die Steuerelemente in "Zu transponierende Variable", um die umstrukturierte Variable in der neuen Datei zu definieren.

So geben Sie eine umstrukturierte Variable an:

- Fügen Sie die Variablen der zu transformierenden Variablengruppe der Liste "Zu transponierende Variable" hinzu. Alle Variablen der Gruppe müssen vom selben Typ sein (numerische oder String-Variablen).

  Dieselbe Variable kann mehrfach in der Variablengruppe vorhanden sein (Variablen werden in der Regel kopiert und nicht aus der Liste der Quellvariablen verschoben). Die Werte werden in der neuen Datei wiederholt.

So geben Sie mehrere umstrukturierte Variablen an:

- Wählen Sie in der Dropdown-Liste "Zielvariable" die erste Zielvariable aus, die Sie definieren möchten.
- Fügen Sie die Variablen der zu transformierenden Variablengruppe der Liste "Zu transponierende Variable" hinzu. Alle Variablen der Gruppe müssen vom selben Typ sein (numerische oder String-Variablen). Eine Variable kann mehr als einmal in der Variablengruppe enthalten sein. (Variablen werden aus der Liste der Quellvariablen nicht verschoben, sondern kopiert, und ihre Werte werden in der neuen Datei wiederholt.)
- Wählen Sie die nächste Zielvariable aus, die Sie definieren möchten, und wiederholen Sie den Auswahlvorgang für alle verfügbaren Zielvariablen.
  - Obwohl eine Variable mehrfach in einer Zielvariablengruppe vorkommen kann, darf dieselbe Variable nicht in mehreren Zielvariablengruppen beinhaltet sein.
  - Jede Liste von Zielvariablengruppen muss die gleiche Anzahl von Variablen enthalten. (Mehrmals aufgeführte Variablen werden in die Zählung einbezogen.)
  - Die Anzahl der Zielvariablengruppen wird von der Anzahl der im vorhergehenden Schritt angegebenen Variablengruppen bestimmt. Die Standardvariablennamen können hier geändert werden, aber um die Anzahl der zu umstrukturierenden Variablengruppen zu ändern, müssen Sie zum vorhergehenden Schritt zurückkehren.
  - Bevor Sie mit dem nächsten Schritt fortfahren können, müssen Sie für alle Zielvariablen eine Variablengruppe definiert haben (durch Auswählen von Variablen in der Quellliste).

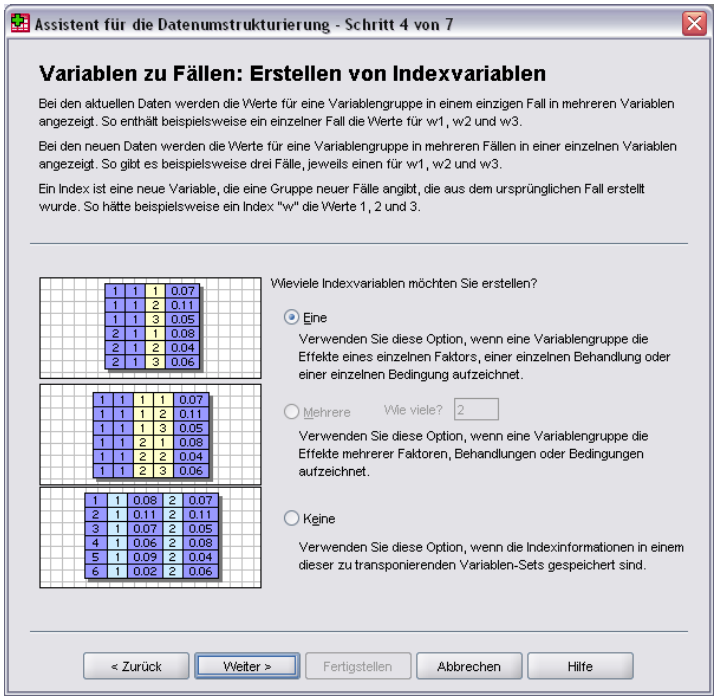
Was soll in die neue Datei kopiert werden? Variablen, die nicht umstrukturiert werden, können in die neue Datei kopiert werden. Die Werte für diese Variablen werden in die neuen Zeilen übertragen. Verschieben Sie die Variablen, die in die neue Datei kopiert werden sollen, in die Liste "Variable(n) mit festem Format".

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Erstellen von Indexvariablen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Variablengruppen in Zeilen umstrukturiert werden sollen.

In diesem Schritt legen Sie fest, ob Indexvariablen erstellt werden sollen. Bei einem Index handelt es sich um eine neue Variable, die eine Zeilengruppe fortlaufend anhand der ursprünglichen Variablen identifiziert, aus der die neue Zeile erstellt wurde.

Abbildung 9-24
Assistent für die Datenumstrukturierung: Erstellen von Indexvariablen, Schritt 4



Wie viele Indexvariablen soll die neue Datei enthalten? Indexvariablen können in Prozeduren als Gruppenvariablen verwendet werden. In den meisten Fällen ist eine einzige Indexvariable ausreichend. Wenn die Variablengruppen in der aktuellen Datei allerdings mehrere Faktorstufen darstellen, sind unter Umständen mehrere Indizes erforderlich.

- **Eine.** Der Assistent erstellt eine einzige Indexvariable.
- **Mehrere.** Der Assistent erstellt mehrere Indizes. Geben Sie die Anzahl der zu erstellenden Indizes ein. Die angegebene Anzahl wirkt sich auf den nächsten Schritt aus, in dem der Assistent automatisch die angegebene Anzahl von Indizes erstellt.
- **Keine.** Wählen Sie diese Option aus, wenn in der neuen Datei keine Indexvariable erstellt werden soll.

Beispiel für einen Index bei der Umstrukturierung von Variablen zu Fällen

Die aktuellen Daten enthalten eine Variablengruppe für die *Breite* und einen Faktor für die *Zeit*. Die Breite wurde dreimal gemessen und in *w1*, *w2* und *w3* aufgezeichnet.

Abbildung 9-25
Aktuelle Daten für einen Index

	subjekt	w1	w2	w3
1	1	6.70	4.30	5.70
2	2	7.10	5.90	5.60

Die Variablengruppe wird nun in eine einzelne Variable für *Breite* umstrukturiert. Zudem wird ein einzelner numerischer Index erstellt. In der folgenden Tabelle werden die neuen Daten abgebildet.

Abbildung 9-26
Neue umstrukturierte Daten mit einem Index

	subjekt	index	breite
1	1	1	6.70
2	1	2	4.30
3	1	3	5.70
4	2	1	7.10
5	2	2	5.90
6	2	3	5.60

Der *Index* beginnt bei 1 und zählt jede Variable in der Gruppe in aufsteigender Reihenfolge. Er beginnt jedesmal neu, wenn in der Originaldatei eine neue Zeile gefunden wird. Der *Index* kann von nun an in Prozeduren verwendet werden, für die eine Gruppenvariable erforderlich ist.

Beispiel für zwei Indizes bei der Umstrukturierung von Variablen zu Fällen

Wenn eine Variablengruppe mehrere Faktoren aufzeichnet, können Sie mehrere Indizes erstellen. Die aktuellen Daten müssen dabei jedoch so angeordnet sein, dass die Stufen des ersten Faktors einen Primärindex bilden, der von allen weiteren Faktorstufen durchlaufen wird. Die aktuelle Datei enthält eine Variablengruppe für die *Breite* und zwei Faktoren, *A* und *B*. Die Daten sind so angeordnet, dass die Stufen von Faktor *B* die Stufen von Faktor *A* durchlaufen.

Abbildung 9-27
Aktuelle Daten für zwei Indizes

	subjekt	w_a1b1	w_a1b2	w_a2b1	w_a2b2
1	1	5.50	6.40	5.80	5.90
2	2	7.40	7.10	5.60	6.70

Die Variablengruppe wird nun in eine einzelne Variable für die *Breite* umstrukturiert, und dabei werden zwei Indizes erstellt. In der folgenden Tabelle werden die neuen Daten abgebildet.

Abbildung 9-28
Neue, umstrukturierte Daten mit zwei Indizes

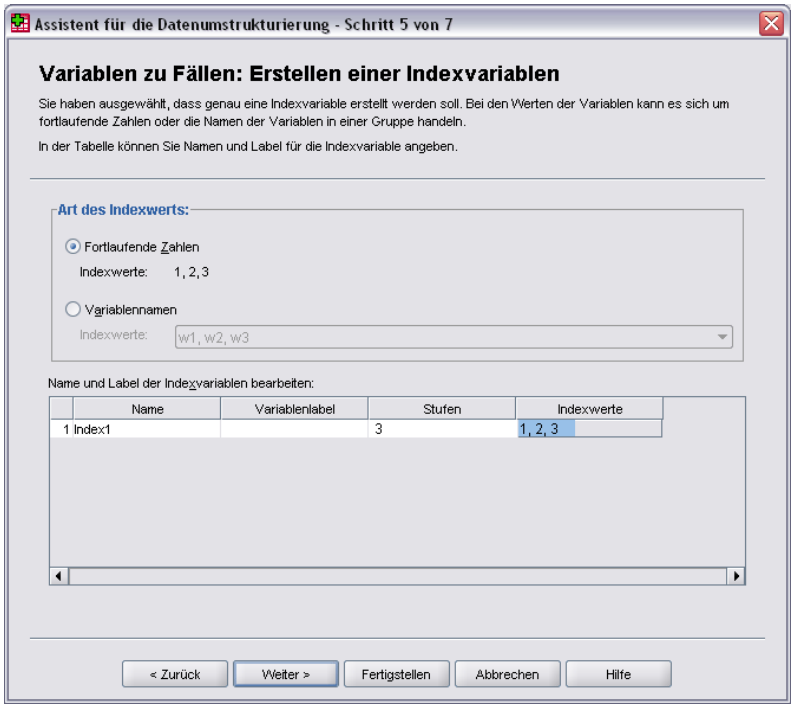
	subjekt	index_a	index_b	breite
1	1	1	1	5.50
2	1	1	2	6.40
3	1	2	1	5.80
4	1	2	2	5.90
5	2	1	1	7.40
6	2	1	2	7.10
7	2	2	1	5.60
8	2	2	2	6.70

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Erstellen einer Indexvariablen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Sie die Umstrukturierung von Variablengruppen in Zeilen und die Erstellung von einer Indexvariablen auswählen.

In diesem Schritt legen Sie die gewünschten Werte für die Indexvariable fest. Bei den Werten kann es sich um fortlaufende Nummern oder um die Namen der Variablen aus der ursprünglichen Variablengruppe handeln. Außerdem können Sie einen Namen und ein Label für die neue Indexvariable angeben.

Abbildung 9-29
Assistent für die Datenumstrukturierung: Erstellen einer Indexvariablen, Schritt 5



Für weitere Informationen siehe Beispiel für einen Index bei der Umstrukturierung von Variablen zu Fällen auf S. 214.

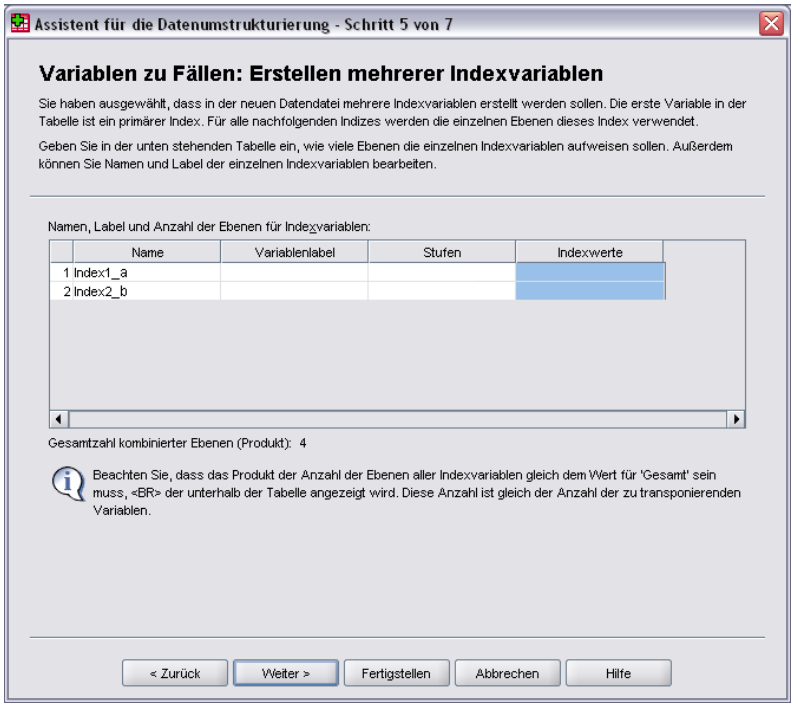
- **Fortlaufende Zahlen.** Der Assistent weist automatisch fortlaufende Nummern als Indexwerte zu.
- **Variablennamen.** Der Assistent verwendet die Namen der ausgewählten Variablengruppe als Indexwerte. Wählen Sie eine Variablengruppe in der Liste aus.
- **Namen und Labels.** Klicken Sie auf eine Zelle, um den vorgegebenen Variablennamen zu ändern und ein beschreibendes Variablenlabel für die Indexvariable einzugeben.

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Erstellen mehrerer Indexvariablen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Variablengruppen in Zeilen umstrukturiert und mehrere Indexvariablen erstellt werden sollen.

In diesem Schritt geben Sie die Anzahl der Stufen für jede Indexvariable an. Außerdem können Sie einen Namen und ein Label für die neue Indexvariable angeben.

Abbildung 9-30
Assistent für die Datenumstrukturierung: Erstellen mehrerer Indexvariablen, Schritt 5



Für weitere Informationen siehe Beispiel für zwei Indizes bei der Umstrukturierung von Variablen zu Fällen auf S. 215.

Wie viele Stufen sind in der aktuellen Datei aufgezeichnet? Überprüfen Sie, wie viele Faktorstufen in den aktuellen Daten aufgezeichnet sind. Eine **Stufe** definiert eine Gruppen von Fällen, die identischen Bedingungen ausgesetzt wurden. Wenn mehrere Faktoren vorhanden sind, müssen die aktuellen Daten so angeordnet sein, dass die Stufen des ersten Faktors einen Primärindex darstellen, der von allen weiteren Faktorstufen durchlaufen wird.

Wie viele Stufen soll die neue Datei enthalten? Geben Sie für jeden Index die Anzahl der Stufen ein. Die Werte für mehrere Indexvariablen sind immer fortlaufende Nummern. Die Werte beginnen bei 1 und werden für jede Stufe erhöht. Die Schrittgröße für die Erhöhung der Indizes ist für den ersten Index am kleinsten und für den letzten Index am größten.

Gesamtzahl kombinierter Ebenen. Es können nur so viele Stufen erstellt werden, wie in den aktuellen Daten vorhanden sind. Da die umstrukturierten Daten eine Zeile pro Behandlungskombination enthalten, überprüft der Assistent die Anzahl der zu erstellenden Stufen. Er vergleicht das Produkt der von Ihnen erstellten Stufen mit der Anzahl der Variablen in der Variablengruppe. Die beiden Werte müssen übereinstimmen.

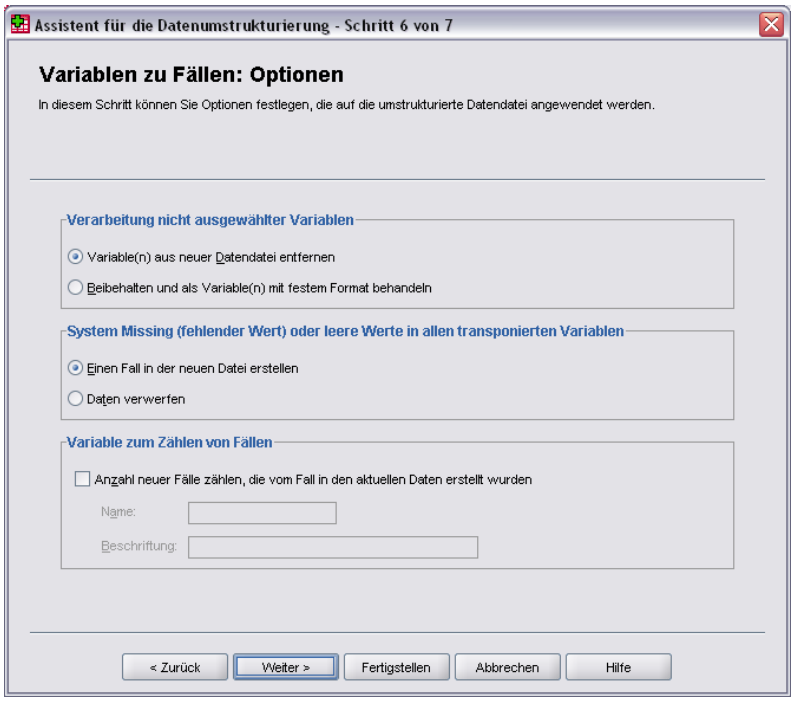
Namen und Labels. Klicken Sie auf eine Zelle, um den vorgegebenen Variablennamen zu ändern und ein beschreibendes Variablenlabel für die Indexvariablen einzugeben.

Assistent für die Datenumstrukturierung (Variablen zu Fällen): Optionen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Variablengruppen in Zeilen umstrukturiert werden sollen.

In diesem Schritt legen Sie die Optionen für die neue umstrukturierte Datei fest.

Abbildung 9-31
Assistent für die Datenumstrukturierung: Optionen, Schritt 6



Sollen nicht ausgewählte Variablen verworfen werden? Im Schritt "Variablen auswählen" (Schritt 3) haben Sie aus den aktuellen Daten die umzustrukturierenden Variablengruppen, die zu kopierenden Variablen und eine Bezeichnervariable ausgewählt. Die Daten der ausgewählten Variablen werden in die neue Datei übertragen. Wenn die aktuellen Daten weitere Variablen enthalten, können Sie festlegen, ob diese verworfen oder beibehalten werden sollen.

Sollen fehlende Daten beibehalten werden? Der Assistent überprüft jede potentiell neue Zeile auf das Vorhandensein von Null-Werten. Ein **Null-Wert** ist ein systemdefinierter fehlender oder leerer Wert. Sie können festlegen, ob Zeilen mit Null-Werten beibehalten oder verworfen werden sollen.

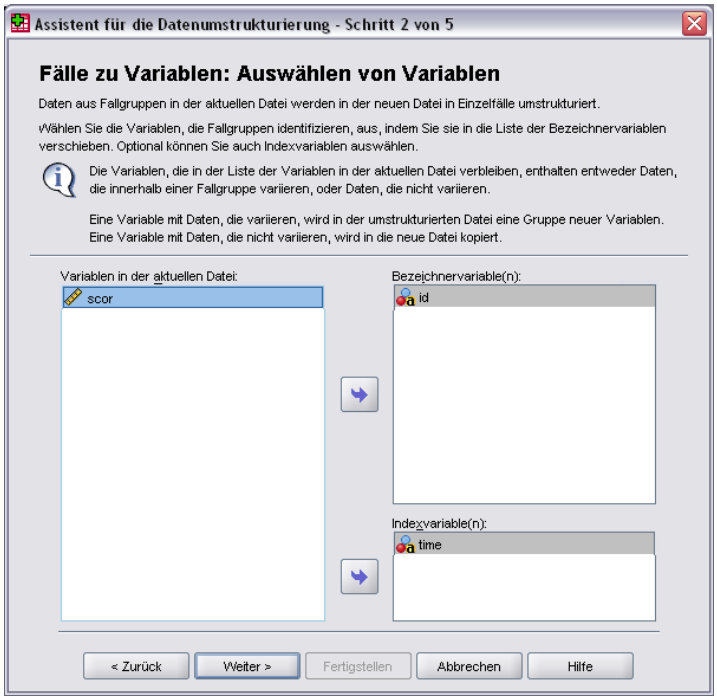
Soll eine Zählvariable erstellt werden? Der Assistent kann in der neuen Datei eine **Zählvariable** erstellen. Diese enthält die Anzahl der neuen Zeilen, die von einer Zeile in den aktuellen Daten erzeugt wurden. Eine Zählvariable kann nützlich sein, wenn Sie die Null-Werte in der neuen Datei verwerfen möchten, da in diesem Fall für eine gegebene Zeile in den aktuellen Daten die Anzahl der neu zu erstellenden Zeilen variieren kann. Klicken Sie auf eine Zelle, um den vorgegebenen Variablennamen zu ändern und ein beschreibendes Variablenlabel für die Zählvariable einzugeben.

Assistent für die Datenumstrukturierung (Fälle zu Variablen): Auswählen von Variablen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Fallgruppen in Spalten umstrukturiert werden sollen.

In diesem Schritt geben Sie an, wie die Variablen der aktuellen Datei in der neuen Datei verwendet werden sollen.

Abbildung 9-32
Assistent für die Datenumstrukturierung: Variablen auswählen, Schritt 2



Wodurch werden Fallgruppen in den aktuellen Daten bezeichnet? Bei einer **Fallgruppe** handelt es sich um eine Gruppe von Zeilen, die aufeinander bezogen sind, weil sie dieselbe Beobachtungseinheit messen, beispielsweise eine Person oder eine Institution. Sie müssen im Assistenten angeben, welche Variablen der aktuellen Datei die Fallgruppen identifizieren, damit die einzelnen Fallgruppen in der neuen Datei jeweils in eine einzelne Zeile konsolidiert werden können. Verschieben Sie die Variablen, die in der aktuellen Datei Fallgruppen identifizieren, in die Liste "Bezeichnervariable(n)". Variablen, die zum Aufteilen der aktuellen Datendatei dienen, werden automatisch zum Identifizieren von Fallgruppen verwendet. Bei jedem Auftreten einer neuen Kombination von Identifizierungswerten wird eine neue Zeile erstellt, sodass die Fälle in der aktuellen Datei nach den Werten der Bezeichnervariablen sortiert sein müssen, und zwar in der Reihenfolge, in der die Variablen in der Liste "Bezeichnervariable(n)" aufgeführt sind. Wenn die aktuelle Datendatei noch nicht sortiert ist, können Sie dies im nächsten Schritt nachholen.

Auf welche Weise sollen die neuen Variablengruppen in der neuen Datei erstellt werden? In den ursprünglichen Daten wird eine Variable in einer einzelnen Spalte aufgeführt. In den neuen Datendatei wird diese Variable in mehreren Spalten aufgelistet. **Indexvariablen** sind Variablen in den aktuellen Daten, die vom Assistenten zum Erstellen der neuen Spalten verwendet werden. Die

umstrukturierten Daten enthalten eine neue Variable für jeden eindeutigen Wert in diesen Spalten. Verschieben Sie die Variablen, die zur Erstellung der neuen Variablengruppen verwendet werden sollen, in die Liste “Indexvariable(n)”. Sie können die neuen Spalten auch nach Indizes ordnen, wenn die entsprechenden Optionen vom Assistenten anzeigt werden.

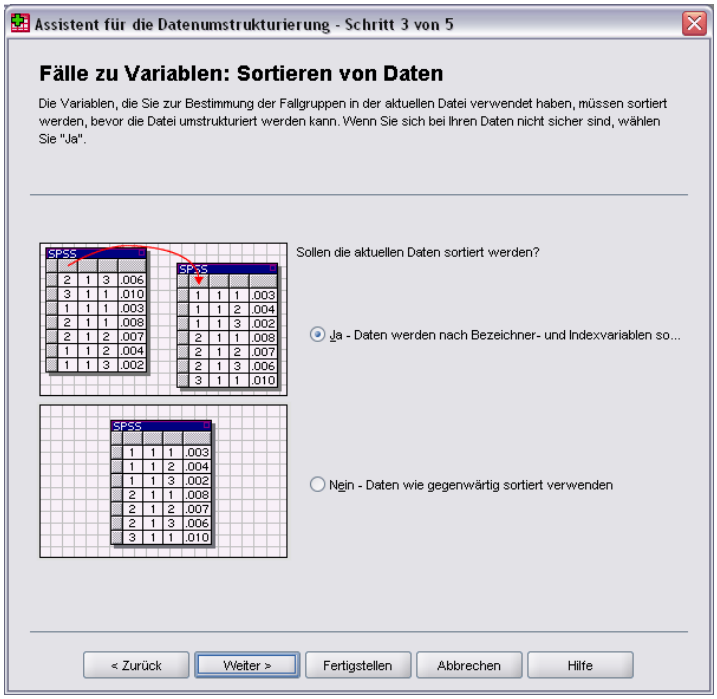
Was passiert mit den anderen Spalten? Der Assistent entscheidet automatisch, was mit den Variablen geschieht, die in der Liste “Aktuelle Datei” verbleiben. Er überprüft jede Variable, um festzustellen, ob die Datenwerte innerhalb einer Fallgruppe variieren. Wenn dies der Fall ist, strukturiert der Assistent die Werte in eine Variablengruppe in der neuen Datei um. Wenn dies nicht der Fall ist, kopiert der Assistent die Werte in die neue Datei.

Assistent für die Datenumstrukturierung (Fälle zu Variablen): Sortieren von Daten

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Fallgruppen in Spalten umstrukturiert werden sollen.

In diesem Schritt legen Sie fest, ob die Daten in der aktuellen Datei vor der Umstrukturierung sortiert werden sollen. Bei jedem Auftreten einer neuen Kombination von Identifizierungswerten wird vom Assistenten eine neue Zeile erstellt. Aus diesem Grund ist es wichtig, dass die Daten nach den Variablen sortiert sind, die die Fallgruppen identifizieren.

Abbildung 9-33
Assistent für die Datenumstrukturierung: Sortieren von Daten, Schritt 3



Wie sind die Zeilen in der aktuellen Datei geordnet? Überprüfen Sie, wie die aktuellen Daten sortiert sind und welche (im vorhergehenden Schritt angegebenen) Variablen zum Identifizieren von Fallgruppen verwendet werden.

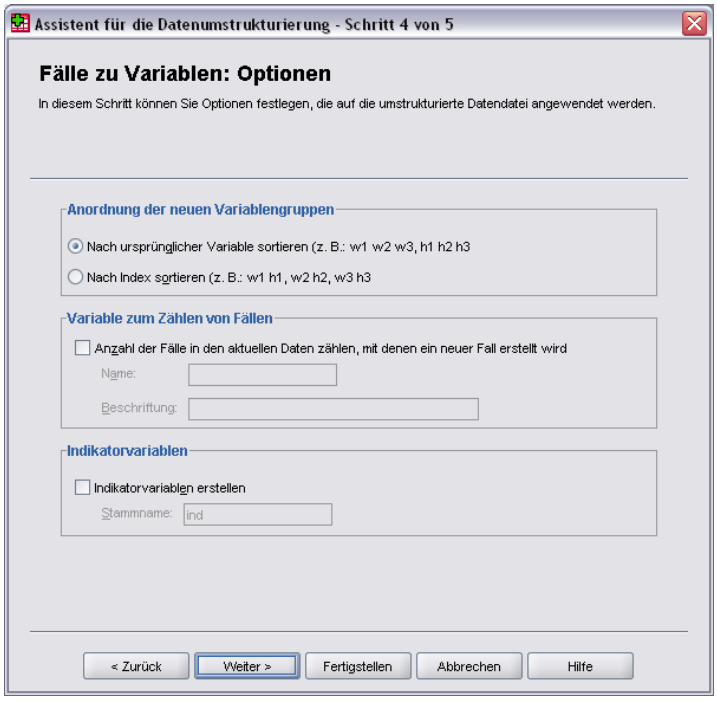
- **Ja.** Der Assistent sortiert die aktuellen Daten automatisch nach den Bezeichnervariablen, und zwar in der Reihenfolge, in der die Variablen im vorhergehenden Schritt in die Liste "Bezeichnervariable(n)" eingetragen wurden. Wählen Sie diese Option aus, wenn die Daten noch nicht nach Bezeichnervariablen sortiert sind oder Sie sich nicht sicher sind. Diese Option erfordert einen zusätzlichen Datendurchlauf, aber sie garantiert, dass die Zeilen für die Umstrukturierung korrekt sortiert sind.
- **Nein.** Der Assistent sortiert die aktuellen Daten nicht. Wählen Sie diese Option aus, wenn Sie sicher sind, dass die aktuellen Daten bereits nach den Variablen sortiert sind, die die Fallgruppen identifizieren.

Assistent für die Datenumstrukturierung (Fälle zu Variablen): Optionen

Anmerkung: Der Assistent zeigt diesen Schritt an, wenn Fallgruppen in Spalten umstrukturiert werden sollen.

In diesem Schritt legen Sie die Optionen für die neue umstrukturierte Datei fest.

Abbildung 9-34
Assistent für die Datenumstrukturierung: Optionen, Schritt 4



Wie sollen die neuen Variablengruppen in der neuen Datei sortiert werden?

- **Nach Variablen.** Der Assistent gruppiert die neuen Variablen nach den ursprünglichen Variablen, aus denen sie erstellt wurden.
- **Nach Index.** Der Assistent gruppiert die Variablen nach den Werten der Indexvariablen.

Beispiel. Die umzustrukturierenden Variablen sind *w* und *h*, und der Index lautet *Monat*:

w h Monat

Die Gruppierung nach Variablen ergibt das folgende Ergebnis:

w.jan w.feb h.jan

Die Gruppierung nach Indizes ergibt das folgende Ergebnis:

w.jan h.jan w.feb

Soll eine Zählvariable erstellt werden? Der Assistent kann in der neuen Datei eine Zählvariable erstellen. Diese Zählvariable enthält die Anzahl der Zeilen in den aktuellen Daten, die zur Erstellung einer Zeile in der neuen Datendatei verwendet wurden.

Sollen Indikatorvariablen erstellt werden? Mithilfe der Indexvariablen kann der Assistent in der neuen Datendatei **Indikatorvariablen** erstellen. Er erstellt für jeden eindeutigen Wert der Indexvariable eine neue Variable. Die Indikatorvariablen geben das Vorhandensein bzw. Nichtvorhandensein eines Werts für einen Fall an. Wenn der Fall einen Wert aufweist, besitzt eine Indikatorvariable den Wert 1, andernfalls besitzt sie den Wert 0.

Beispiel. Die Indexvariable ist *Produkt*. Sie dient zum Aufzeichnen der von einem Kunden erworbenen Produkte. Die ursprünglichen Daten lauten wie folgt:

Kunde	Produkt
1	Huhn
1	Eier
2	Eier
3	Huhn

Beim Erstellen einer Indikatorvariable entsteht eine neue Variable für jeden eindeutigen Wert von *Produkt*. Die umstrukturierten Daten lauten wie folgt:

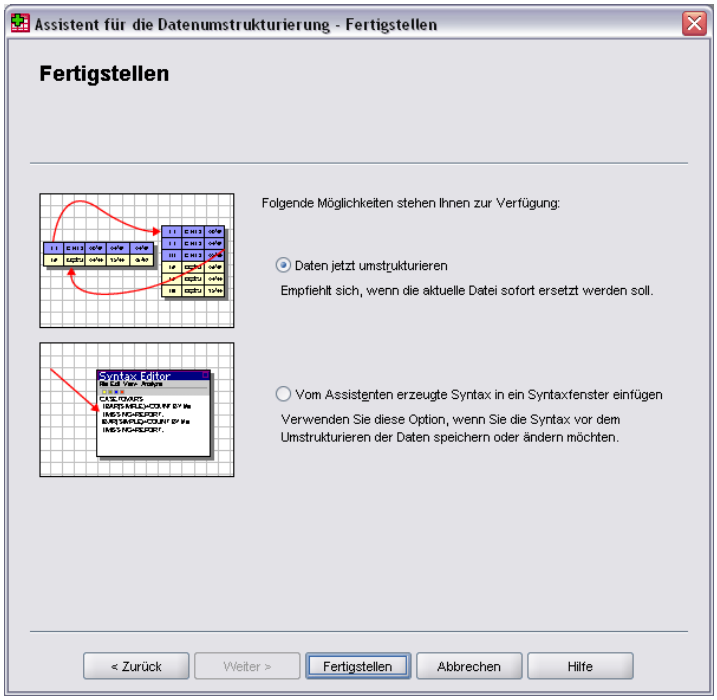
Kunde	indHuhn	indEier
1	1	1
2	0	1
3	1	0

In diesem Beispiel können die umstrukturierten Daten zum Ermitteln der Häufigkeiten für die von Kunden erworbenen Produkte verwendet werden.

Assistent für die Datenumstrukturierung: Fertig stellen

Dies ist der letzte Schritt im Assistenten für die Datenumstrukturierung. Geben Sie an, welcher Vorgang mit den von Ihnen getroffenen Angaben ausgeführt werden soll.

Abbildung 9-35
Assistent für die Datenumstrukturierung: Fertigstellen



- **Daten jetzt umstrukturieren.** Der Assistent erstellt die neue umstrukturierte Datei. Wählen Sie diese Option aus, wenn die aktuelle Datei sofort ersetzt werden soll.

Anmerkung: Wenn die ursprünglichen Daten gewichtet sind, sind auch die neuen Daten gewichtet, es sei denn, die für die Gewichtung verwendete Variable wird in der neuen Datei umstrukturiert oder verworfen.

- **Vom Assistenten erzeugte Syntax in ein Syntaxfenster einfügen**. Der Assistent fügt die von ihm erstellte Syntax in ein Syntax-Fenster ein. Wählen Sie diese Option aus, wenn die aktuelle Datei noch nicht ersetzt werden soll oder wenn die Syntax noch weitergehend bearbeitet bzw. zur späteren Verwendung gespeichert werden soll.

Kapitel
10

Arbeiten mit Ausgaben

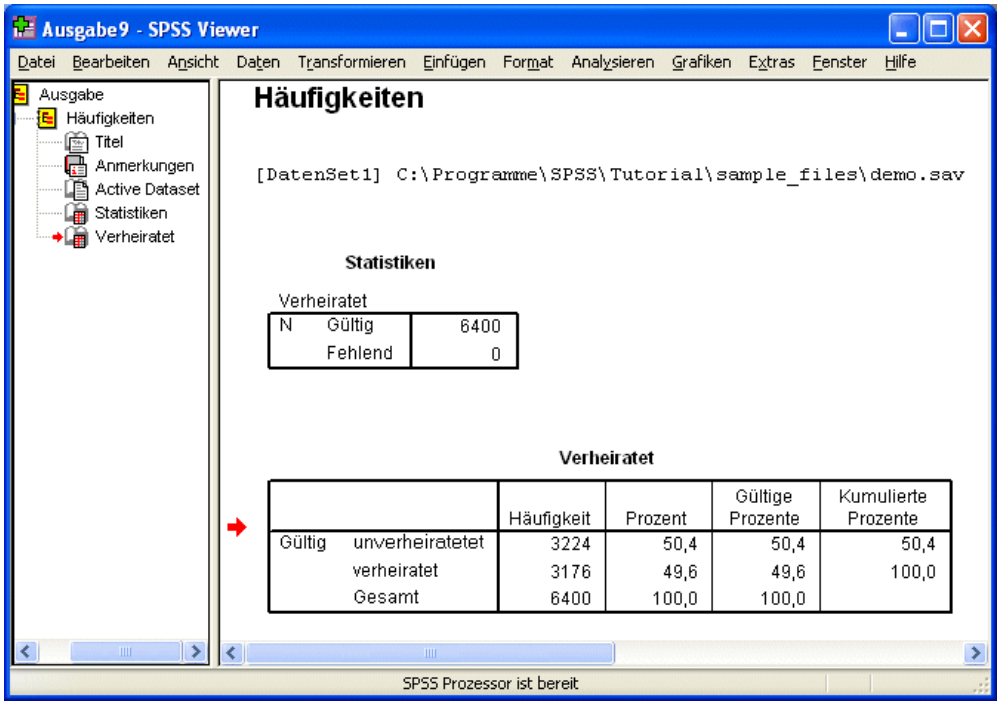
Nach dem Ausführen einer Prozedur werden die Ergebnisse in einem Fenster angezeigt, das "Viewer" heißt. In diesem Fenster können Sie problemlos zwischen den verschiedenen Teilen der Ausgabe wechseln. Außerdem können Sie die Ausgaben bearbeiten und so Dokumente erstellen, die genau die gewünschten Ausgaben enthalten.

Viewer

Die Ergebnisse werden im Viewer angezeigt. Sie können den Viewer für folgende Vorgänge verwenden:

- Durchsuchen der Ergebnisse
- Ein- und Ausblenden von ausgewählten Tabellen und Diagrammen
- Ändern der Anzeigereihenfolge der Ergebnisse durch Verschieben ausgewählter Objekte
- Verschieben von Objekten zwischen dem Viewer und anderen Anwendungen

Abbildung 10-1
Viewer



Der Viewer ist in zwei Fensterbereiche aufgeteilt:

- Der linke Fensterbereich enthält eine Gliederungsansicht des Inhalts.
- Der rechte Fensterbereich enthält Statistiktabellen, Diagramme und Textausgabe.

Sie können auf ein Element in der Gliederung klicken, um direkt zur zugehörigen Tabelle bzw. dem zugehörigen Diagramm zu wechseln. Wenn Sie die Breite des Gliederungsfensters ändern möchten, können Sie auf dessen rechten Rahmen klicken und ihn mit gedrückter Maustaste auf die gewünschte Breite ziehen.

Ein- und Ausblenden von Ergebnissen

Im Viewer können Sie ausgewählte Tabellen oder alle Ergebnisse einer Prozedur ein- und ausblenden. Dies ist nützlich, wenn Sie möchten, dass im Inhaltsfenster weniger angezeigt wird.

So blenden Sie Tabellen und Diagramme aus:

▶ Doppelklicken Sie im Gliederungsfenster des Viewers auf das Buchsymbol des Objekts.

oder

▶ Klicken Sie einmal auf das Objekt, um es auszuwählen.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Ansicht
  Ausblenden

oder

▶ Klicken Sie in der Gliederungs-Symbolleiste auf die Schaltfläche “Ausblenden”, die durch ein geschlossenes Buch dargestellt ist.

Die Schaltfläche “Einblenden”, die durch ein geöffnetes Buch dargestellt ist, wird zu einer aktiven Schaltfläche. Dadurch wird angezeigt, dass das Objekt jetzt ausgeblendet ist.

So blenden Sie die Ergebnisse einer Prozedur aus:

▶ Klicken Sie im Gliederungsfenster auf das Kästchen neben dem Namen der Prozedur.

Dadurch werden alle Ergebnisse der Prozedur ausgeblendet und die Gliederungsansicht reduziert.

Verschieben, Löschen und Kopieren von Ausgaben

Sie können die Ergebnisse neu anordnen, indem Sie einzelne Objekte oder ganze Objektgruppen kopieren, verschieben oder löschen.

So verschieben Sie Ausgaben im Viewer:

▶ Wählen Sie die Objekte im Gliederungs- oder Inhaltsfenster aus.

▶ Ziehen Sie die Objekte auf einen anderen Speicherort und legen Sie sie dort ab.

So löschen Sie Ausgaben im Viewer:

- ▶ Wählen Sie die Objekte im Gliederungs- oder Inhaltsfenster aus.
- ▶ Drücken Sie die Entf-Taste.

  oder

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Löschen

Ändern der anfänglichen Ausrichtung

In der Standardeinstellung sind alle Ergebnisse linksbündig ausgerichtet. So ändern Sie die ursprüngliche Ausrichtung neuer Ausgabeobjekte:

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Optionen

- ▶ Klicken Sie auf die Registerkarte Viewer.
- ▶ Wählen Sie in der Gruppe "Anfänglicher Ausgabestatus" den gewünschten Objekttyp aus (z. B. Pivot-Tabelle, Diagramm, Textausgabe).
- ▶ Wählen Sie die gewünschte Ausrichtungsoption aus.

Ändern der Ausrichtung von Ausgabeobjekten

- ▶ Wählen Sie im Gliederungs- bzw. Inhaltsbereich die Elemente aus, die sie ausrichten möchten.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

  Format
  Linksbündig

  oder

  Format
  Mitte

  oder

  Format
  Rechtsbündig

Gliederung des Viewers

Im Gliederungsfenster wird eine Inhaltsangabe des Viewer-Dokuments angezeigt. Hier können Sie die Ergebnisse durchblättern und festlegen, welche Objekte angezeigt werden. Die meisten Aktionen im Gliederungsfenster wirken sich auch auf das Inhaltsfenster aus.

- Beim Auswählen eines Objekts im Gliederungsfenster wird das entsprechende Objekt im Inhaltsfenster angezeigt.

- Beim Verschieben eines Objekts im Gliederungsfenster wird das entsprechende Objekt im Inhaltsfenster ebenfalls verschoben.
- Beim Reduzieren der Gliederungsansicht werden die Ergebnisse aller Objekte in den reduzierten Ebenen ausgeblendet.

Einstellen der Anzeige für die Gliederung. Zum Einstellen der Anzeige für die Gliederung stehen Ihnen die folgenden Möglichkeiten zur Verfügung:

- Erweitern und Reduzieren der Gliederungsansicht
- Ändern der Gliederungsebene von ausgewählten Objekten
- Ändern der Größe von Objekten im Gliederungsfenster
- Ändern der in der Gliederung verwendeten Schriftart

So reduzieren und erweitern Sie die Gliederungsansicht:

▶ Klicken Sie auf das Kästchen links neben dem Gliederungsobjekt, das Sie reduzieren oder erweitern möchten.

oder

▶ Klicken Sie in der Gliederung auf das Objekt.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Ansicht
  Reduzieren

oder

Ansicht
  Erweitern

So ändern Sie die Gliederungsebene:

▶ Klicken Sie im Gliederungsfenster auf das Objekt.

▶ Klicken Sie in der Gliederungs-Symbolleiste auf den nach links zeigenden Pfeil, um das Objekt heraufzustufen (Verschieben des Objekts nach links).

oder

Klicken Sie in der Gliederungs-Symbolleiste auf den nach rechts zeigenden Pfeil, um das Objekt herabzustufen (Verschieben des Objekts nach rechts).

oder

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Gliederung
    Heraufstufen

oder

Bearbeiten
  Gliederung
    Herabstufen

Das Ändern der Gliederungsebene ist besonders nach dem Verschieben von Objekten im Gliederungsfenster nützlich. Beim Verschieben der Objekte kann sich die Gliederungsebene der Objekte ändern. In diesem Fall können Sie die Pfeil-Schaltflächen auf der Gliederungs-Symbolleiste verwenden, um die ursprüngliche Gliederungsebene wiederherzustellen.

So ändern Sie die Größe von Objekten in der Gliederung:

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Ansicht
  Größe der Gliederung
- ▶ Wählen Sie die Größe der Gliederung (Klein, Mittel oder Groß).

So ändern Sie die Schriftart in der Gliederung:

- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Ansicht
  Schriftart für Gliederung...
- ▶ Wählen Sie eine Schriftart aus.

Einfügen von Objekten im Viewer

Sie können im Viewer Objekte einfügen, beispielsweise Titel, neue Texte, Diagramme oder Objekte aus anderen Anwendungen.

So fügen Sie einen neuen Titel oder ein neues Textobjekt ein:

Sie können im Viewer Textobjekte einfügen, die nicht mit einer Tabelle oder einem Diagramm verbunden sind.

- ▶ Klicken Sie auf die Tabelle, das Diagramm oder das Objekt, nach der bzw. dem der Titel oder Text eingefügt werden soll.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Einfügen
  Neuer Titel

  oder

  Einfügen
  Neuer Text
- ▶ Doppelklicken Sie auf das neue Objekt.
- ▶ Geben Sie den gewünschten Text ein.

So fügen Sie eine Textdatei hinzu:

- Klicken Sie im Gliederungs- oder Inhaltsfenster des Viewers auf die Tabelle, das Diagramm oder das Objekt, nach der bzw. dem der Text eingefügt werden soll.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Einfügen
  Textdatei...
- Wählen Sie eine Textdatei aus.

  Doppelklicken Sie auf die Textdatei, um diese zu bearbeiten.

Einfügen von Objekten im Viewer

Sie können Objekte aus anderen Anwendungen in den Viewer einfügen. Verwenden Sie hierfür den Befehl “Einfügen nach” oder “Inhalte einfügen”. Dabei wird das neue Objekt nach dem aktuell ausgewählten Objekt im Viewer eingefügt. Verwenden Sie den Befehl “Inhalte einfügen”, wenn Sie das Format für das einzufügende Objekt auswählen möchten.

Suchen und Ersetzen von Informationen im Viewer

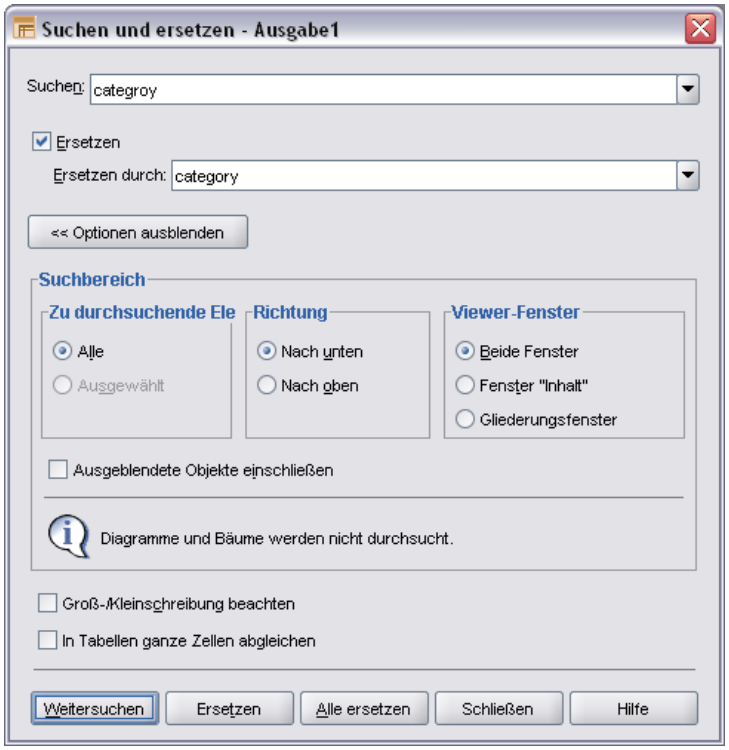
- Um Informationen im Viewer zu suchen bzw. zu ersetzen, wählen Sie folgende Optionen aus den Menüs aus:

  Bearbeiten
  Suchen

  oder

  Bearbeiten
  Ersetzen

Abbildung 10-2
Dialogfeld "Suchen und ersetzen"



Sie können "Suchen und Ersetzen" für folgende Vorgänge verwenden:

- Durchsuchen des gesamten Dokuments bzw. der ausgewählten Objekte.
- Suchen nach unten bzw. oben von der aktuellen Position.
- Durchsuchen beider Fensterbereiche oder Einschränkung der Suche auf den Inhalt des Gliederungsfensters.
- Suchen nach ausgeblendeten Objekten. Dazu gehören alle Objekte, die im Inhaltsfenster ausgeblendet sind (z. B. Anmerkungstabellen, die standardmäßig ausgeblendet sind) sowie ausgeblendete Zeilen und Spalten in Pivot-Tabellen.
- Einschränken der Suchkriterien auf Abhängige Variable, die auch hinsichtlich der Groß- und Kleinschreibung übereinstimmen.
- Einschränken der Suchkriterien in Pivot-Tabellen auf Abhängige Variable, bei denen der gesamte Zelleninhalt übereinstimmt.

Ausgeblendete Objekte und Schichten von Pivot-Tabellen

- Schichten unterhalb der sichtbaren Schicht einer mehrdimensionalen Pivot-Tabelle gelten nicht als ausgeblendet und sind auch dann in den Suchbereich eingeschlossen, wenn ausgeblendete Objekte nicht in die Suche mit aufgenommen wurden.
- Zu den ausgeblendeten Objekten gehören ausgeblendete Objekte im Inhaltsfenster (Objekte, die im Gliederungsfenster mit einem geschlossenen Buch dargestellt werden oder in reduzierten Blöcken des Gliederungsfensters enthalten sind) und Zeilen und Spalten in

Pivot-Tabellen, die entweder standardmäßig ausgeblendet sind (z. B. sind leere Zeilen und Spalten standardmäßig ausgeblendet) oder manuell (durch Bearbeiten der Tabelle und selektives Ausblenden bestimmter Zeilen bzw. Spalten) ausgeblendet sind. Ausgeblendete Objekte werden nur dann in die Suche mit aufgenommen, wenn Sie explizit die Option Ausgeblendete Objekte einschließen auswählen.

- In beiden Fällen wird das ausgeblendete bzw. unsichtbare Element, das den Suchtext bzw. -wert enthält beim Auffinden angezeigt, das Element wird jedoch anschließend wieder in den ursprünglichen Zustand zurückversetzt.

Kopieren von Ausgaben in andere Anwendungen

Ausgabeobjekte können kopiert und in andere Anwendungen, beispielsweise in Textverarbeitungs- oder Tabellenkalkulationsprogramme eingefügt werden. Die Ausgaben können in verschiedenen Formaten eingefügt werden. Je nach Zielanwendung stehen einige oder alle der folgenden Formate zur Verfügung:

Grafik (Metadatei). Pivot-Tabellen, Textausgaben und Diagramme können als Grafiken im Metadatei-Format eingefügt werden. Die Größe der Grafiken kann in anderen Anwendungen geändert werden, und unter Umständen können die Grafiken in begrenztem Umfang auch mit den Funktionen der anderen Anwendungen bearbeitet werden. Bei Pivot-Tabellen, die als Bilder eingefügt wurden, bleiben alle Rahmen und Schriftartenmerkmale erhalten. Dieses Format ist nur unter Windows-Betriebssystemen verfügbar.

RTF (Rich Text Format). Pivot-Tabellen können in andere Anwendungen im RTF-Format eingefügt werden. In den meisten Anwendungen wird die Pivot-Tabelle dabei als Tabelle eingefügt, die dann in der anderen Anwendung bearbeitet werden kann.

Anmerkung: Besonders breite Tabellen werden von Microsoft Word möglicherweise nicht ordnungsgemäß angezeigt.

Bitmap. Pivot-Tabellen und Diagramme können in andere Anwendungen als Bitmaps eingefügt werden.

BIFF. Der Inhalt von Tabellen kann in eine Tabellenkalkulation eingefügt werden, wobei die numerische Genauigkeit erhalten bleibt.

Text. Der Inhalt von Tabellen kann als Text kopiert und in andere Anwendungen eingefügt werden. Dies kann bei Anwendungen wie E-Mail-Programmen nützlich sein, bei denen mit der Anwendung nur Text verarbeitet oder übertragen werden kann.

Wenn die Zielanwendung mehrere verfügbare Formate unterstützt, verfügt sie möglicherweise über eine Menüoption vom Typ “Inhalte einfügen”, mit der Sie das Format auswählen können, oder es wird automatisch eine Liste der verfügbaren Formate angezeigt.

So kopieren Sie Ausgabeobjekte und fügen diese in eine andere Anwendung ein:

- ▶ Wählen Sie die Objekte im Gliederungs- oder Inhaltsfenster des Viewers aus.

▶ Wählen Sie die folgenden Befehle aus den Menüs des Viewers aus:

Bearbeiten
  Kopieren

▶ Wählen Sie die folgenden Befehle aus den Menüs der Ziel-Anwendung aus:

Bearbeiten
  Einfügen

oder

Bearbeiten
  Inhalte einfügen...

Einfügen. Ausgaben werden in einer Reihe von Formaten in die Zwischenablage kopiert. Jede Anwendung ermittelt das "beste" Format für das Einfügen.

Inhalte einfügen. Ergebnisse werden in einer Reihe von Formaten in die Zwischenablage kopiert. Mit dem Befehl "Inhalte einfügen" können Sie das gewünschte Format aus der Liste der verfügbaren Formate in der Ziel-Anwendung auswählen.

Ausgabe exportieren

Mit "Ausgabe exportieren" wird die Viewer-Ausgabe im HTML-,Text-, Word/RTF-, Excel-, PowerPoint- (PowerPoint 97 oder höher erforderlich) oder PDF-Format gespeichert. Außerdem ist der Export von Diagrammen in mehreren verschiedenen Grafikformaten möglich.

Anmerkung: Der Export nach PowerPoint ist nur unter Windows-Betriebssystemen und nicht in der Studentenversion verfügbar.

So exportieren Sie Ausgaben:

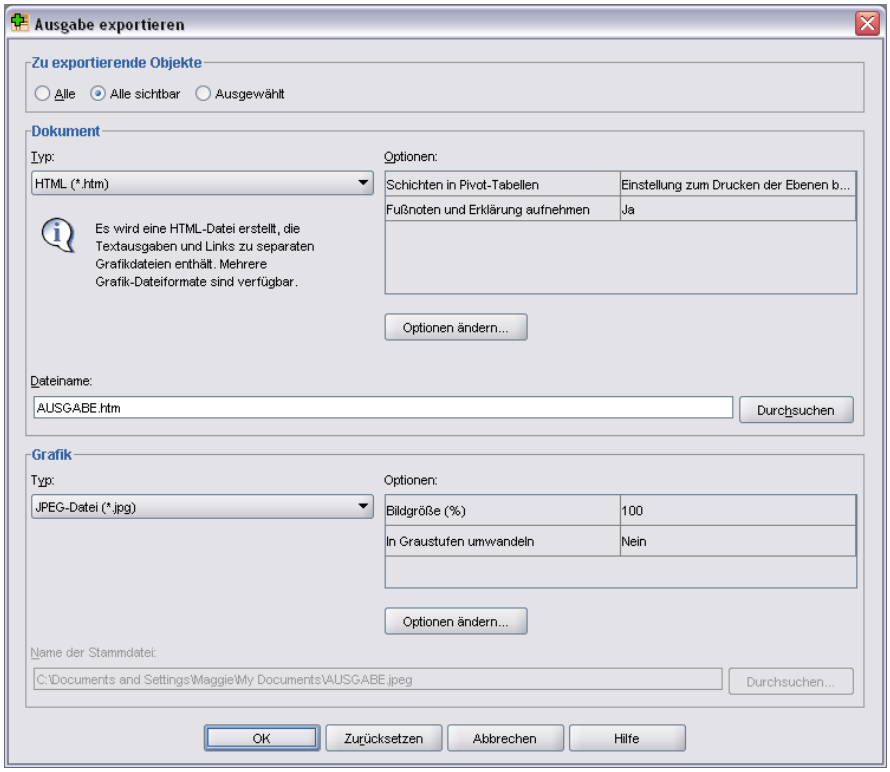
▶ Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Exportieren...

▶ Geben Sie einen Dateinamen oder ein Präfix für Diagramme ein und wählen Sie ein Exportformat aus.

Abbildung 10-3
Dialogfeld "Ausgabe exportieren"



Zu exportierende Objekte. Sie können alle Objekte im Viewer, alle sichtbaren Objekte oder nur ausgewählte Objekte exportieren.

Dokumenttyp. Die folgenden Optionen sind verfügbar:

- **Word/RTF (*.doc).** Pivot-Tabellen werden mit sämtlichen Formatierungsattributen wie Zellenrahmen, Schriftarten und Hintergrundfarben, als Word-Tabellen exportiert. Textausgaben werden als formatierter RTF-Text exportiert. Unter Windows-Betriebssytemen werden Diagramme im Format EMF (Enhanced Metafile) in das Dokument aufgenommen. Unter anderen Betriebssystemen werden Diagramme im Format PNG aufgenommen. Textausgaben werden immer mit einem nicht proportionalen Zeichensatz (mit festem Abstand) angezeigt und mit denselben Schriftartenattributen exportiert. Für die richtige Ausrichtung von durch Leerzeichen getrennten Textausgaben ist ein nicht proportionaler Zeichensatz (mit festem Abstand) erforderlich.

  Anmerkung: Besonders breite Tabellen werden von Microsoft Word möglicherweise nicht ordnungsgemäß angezeigt.

- **Excel (*.xls).** Die Zeilen, Spalten und Zellen von Pivot-Tabellen werden mit sämtlichen Formatierungsattributen wie Zellenrahmen, Schriftarten, Hintergrundfarben als Excel-Zeilen, -Spalten und -Zellen exportiert. Textausgaben werden mit allen Schriftartattributen exportiert. Jede Zeile in der Textausgabe entspricht einer Zeile in der Excel-Datei, wobei der gesamte Inhalt der Zeile in einer einzelnen Zelle enthalten ist. Diagramme werden nicht mit aufgenommen.

- **HTML (*.htm).** Pivot-Tabellen werden als HTML-Tabellen exportiert. Textausgaben werden als vorformatierter HTML-Text exportiert. Diagramme werden als Verweis eingebettet. Daher sollten Sie Diagramme in einem für die Aufnahme in HTML-Dokumente geeigneten Format exportieren (z. B. PNG oder JPEG).
- **Portable Document Format (*.pdf).** Alle Ausgaben werden so exportiert, wie sie in der Druckvorschau/Seitenansicht angezeigt werden. Alle Formatierungsattribute bleiben erhalten.
- **PowerPoint file (*.ppt).** Pivot-Tabellen werden als Word-Dateien exportiert und sind auf separaten Folien in der PowerPoint-Datei eingebettet (je eine Pivot-Tabelle auf einer Folie). Sämtliche Formatierungsattribute der Pivot-Tabelle (z. B. Zellenrahmen, Schriftarten und Hintergrundfarben) werden beibehalten. Diagramme werden im Format TIFF exportiert. Textausgaben sind nicht eingeschlossen.

  Anmerkung: Der Export nach PowerPoint ist nur unter Windows-Betriebssystemen und nicht in der Studentenversion verfügbar.
- **Text (*.txt).** Zu den Textausgabeformaten gehören einfacher Text, UTF-8 und UTF-16. Pivot-Tabellen können als durch Tabulatoren getrennter Text oder als durch Leerzeichen getrennter Text exportiert werden. Alle Textausgaben werden in durch Leerzeichen getrenntem Format exportiert. Bei Diagrammen wird in der Textdatei für jedes Diagramm eine Zeile mit der Angabe des Dateinamens für das exportierte Diagramm eingefügt.
- **Ohne (nur Grafiken).** Folgende Exportformate sind verfügbar: EPS, JPEG, TIFF, PNG und BMP. Unter Windows-Betriebssystemen ist außerdem das Format EMF (Enhanced Metafile, erweiterte Metadatei) verfügbar.

Ausgabeverwaltungssystem Sie können auch automatisch alle Ausgaben oder vom Benutzer festgelegte Ausgabetypen als Datendateien im Text-, HTML-, XML- oder SPSS-Format exportieren. Für weitere Informationen siehe Ausgabeverwaltungssystem (OMS) in Kapitel 48 auf S. 523.

Optionen für HTML, Word/RTF und Excel

Folgende Optionen stehen für den Export von Pivot-Tabellen nach Word/RTF, Excel und HTML zur Verfügung:

- **Schichten in Pivot-Tabellen.** Standardmäßig richtet sich die Aufnahme bzw. der Ausschluss von Pivot-Tabellen-Schichten nach den Tabelleneigenschaften der einzelnen Pivot-Tabellen. Sie können diese Einstellung außer Kraft setzen und alle Schichten aufnehmen oder alle Schichten mit Ausnahme der aktuell sichtbaren Schicht ausschließen. Für weitere Informationen siehe Tabelleneigenschaften: Drucken in Kapitel 11 auf S. 259.
- **Fußnoten und Erklärungen aufnehmen.** Dient zur Festlegung der Aufnahme bzw. des Ausschlusses aller Fußnoten und Erklärungen von Pivot-Tabellen.

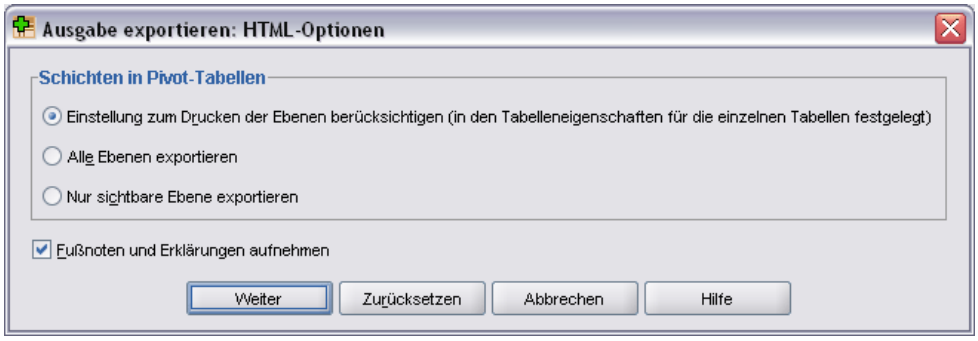
Anmerkung: Bei HTML können Sie außerdem das Bilddateiformat für exportierte Diagramme festlegen. Für weitere Informationen siehe Optionen zum Exportieren von Diagrammen auf S. 238.

So legen Sie die Exportoptionen für HTML, Word/RTF und Excel fest:

▶ Wählen Sie als Exportformat HTML, Word/RTF oder Excel aus.

▶ Klicken Sie auf Optionen ändern.

Abbildung 10-4
Ausgabeexportoptionen für HTML, Word/RTF und Excel



PowerPoint-Optionen

Die folgenden Optionen sind für PowerPoint verfügbar:

- **Schichten in Pivot-Tabellen.** Standardmäßig richtet sich die Aufnahme bzw. der Ausschluss von Pivot-Tabellen-Schichten nach den Tabelleneigenschaften der einzelnen Pivot-Tabellen. Sie können diese Einstellung außer Kraft setzen und alle Schichten aufnehmen oder alle Schichten mit Ausnahme der aktuell sichtbaren Schicht ausschließen. Für weitere Informationen siehe Tabelleneigenschaften: Drucken in Kapitel 11 auf S. 259.
- **Fußnoten und Erklärungen aufnehmen.** Dient zur Festlegung der Aufnahme bzw. des Ausschlusses aller Fußnoten und Erklärungen von Pivot-Tabellen.
- **Einträge in der Viewer-Gliederung als Folientitel verwenden.** Fügt auf jeder beim Export erzeugten Folie einen Titel ein. Jede Folie enthält ein einzelnes Element, das aus dem Viewer exportiert wurde. Der Titel wird aus dem Gliederungseintrag für das Element im Gliederungsfenster des Viewers gebildet.

So legen Sie Exportoptionen für PowerPoint fest:

▶ Wählen Sie PowerPoint als Exportformat.

▶ Klicken Sie auf Optionen ändern.

Anmerkung: Der Export nach PowerPoint ist nur unter Windows-Betriebssystemen verfügbar.

PDF-Optionen

Die folgenden Optionen sind für PDF verfügbar:

Lesezeichen einbetten. Mit dieser Option werden Lesezeichen in das PDF-Dokument aufgenommen, die den Einträgen in der Viewer-Gliederung entsprechen. Wie das Viewer-Gliederungsfenster können auch Lesezeichen die Navigation in Dokumenten mit einer Vielzahl an Ausgabeobjekten erheblich erleichtern.

Schriftarten einbetten. Durch das Einbetten von Schriftarten wird sichergestellt, dass das PDF-Dokument auf allen Computern gleich dargestellt wird. Anderenfalls kann es, wenn im Dokument verwendete Schriftarten auf dem Computer, der zur Anzeige (oder zum Drucken) des

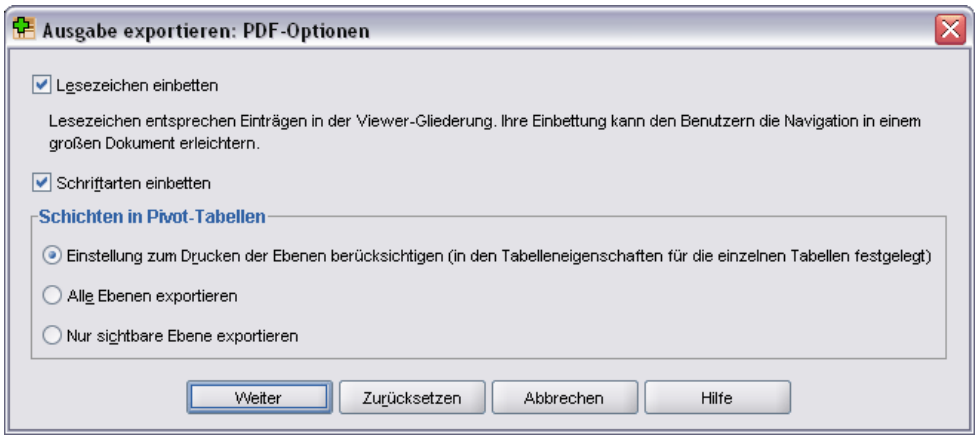
PDF-Dokuments verwendet wird, nicht zur Verfügung stehen, durch Schriftartenersetzung zu suboptimalen Ergebnissen kommen.

Schichten in Pivot-Tabellen. Standardmäßig richtet sich die Aufnahme bzw. der Ausschluss von Pivot-Tabellen-Schichten nach den Tabelleneigenschaften der einzelnen Pivot-Tabellen. Sie können diese Einstellung außer Kraft setzen und alle Schichten aufnehmen oder alle Schichten mit Ausnahme der aktuell sichtbaren Schicht ausschließen. Für weitere Informationen siehe Tabelleneigenschaften: Drucken in Kapitel 11 auf S. 259.

So legen Sie PDF-Exportoptionen fest:

- Wählen Sie Portable Document Format als Exportformat.
- Klicken Sie auf Optionen ändern.

Abbildung 10-5
Dialogfeld "PDF-Optionen"



Andere Einstellungen, die die PDF-Ausgabe beeinflussen

Seite einrichten/Seitenattribute. Seitengröße, Ausrichtung, Ränder, Inhalt und Anzeige von Kopf- und Fußzeilen sowie die Größe des gedruckten Diagramms in PDF-Dokumenten werden über die Optionen für die Seiteneinrichtung und Seitenattribute gesteuert.

Tabelleneigenschaften/Tabellenvorlagen. Die Skalierung breiter und/oder langer Tabellen und der Druck von Tabellenschichten werden durch die Tabelleneigenschaften für die einzelnen Tabellen gesteuert. Diese Eigenschaften können auch unter "Tabellenvorlagen" gespeichert werden. Für weitere Informationen siehe Tabelleneigenschaften: Drucken in Kapitel 11 auf S. 259.

Standarddrucker/aktueller Drucker. Die Auflösung (DPI) des PDF-Dokuments ist die aktuelle Auflösungseinstellung für den Standarddrucker bzw. den aktuell ausgewählten Drucker (kann über "Seite einrichten" geändert werden). Die maximale Auflösung beträgt 1200 DPI. Wenn eine höhere Druckerauflösung eingestellt ist, wird für das PDF-Dokument eine Auflösung von 1200 DPI verwendet.

Anmerkung: Dokumente mit höherer Auflösung können beim Drucken auf Druckern mit niedrigerer Auflösung zu schlechten Ergebnissen führen.

Text: Optionen

Die folgenden Optionen sind für den Textexport verfügbar:

Pivot-Tabellen-Format. Pivot-Tabellen können als durch Tabulatoren getrennter Text oder als durch Leerzeichen getrennter Text exportiert werden. Beim leerzeichengetrennten Format können Sie außerdem folgende Eigenschaften festlegen:

- **Spaltenbreite.** Mit Automatisch anpassen werden keine Spalteninhalte umgebrochen und jede Spalte ist so breit, wie das längste Label bzw. der längste Wert in der Spalte. Mit Benutzerdefiniert wird eine maximale Spaltenbreite festgelegt, die für alle Spalten in der Tabelle gilt. Bei Werten, die breiter sind, wird ein Zeilenumbruch durchgeführt, sodass sie sich auch auf die nächste Zeile in der betreffenden Spalte erstrecken.
- **Zeilen-/Spaltenbegrenzungszeichen.** Legt die Zeichen fest, die für Zeilen- und Spaltenrahmen verwendet werden. Um die Anzeige von Zeilen- und Spaltenrahmen zu unterdrücken, geben Sie als Werte Leerzeichen ein.

Schichten in Pivot-Tabellen. Standardmäßig richtet sich die Aufnahme bzw. der Ausschluss von Pivot-Tabellen-Schichten nach den Tabelleneigenschaften der einzelnen Pivot-Tabellen. Sie können diese Einstellung außer Kraft setzen und alle Schichten aufnehmen oder alle Schichten mit Ausnahme der aktuell sichtbaren Schicht ausschließen. Für weitere Informationen siehe Tabelleneigenschaften: Drucken in Kapitel 11 auf S. 259.

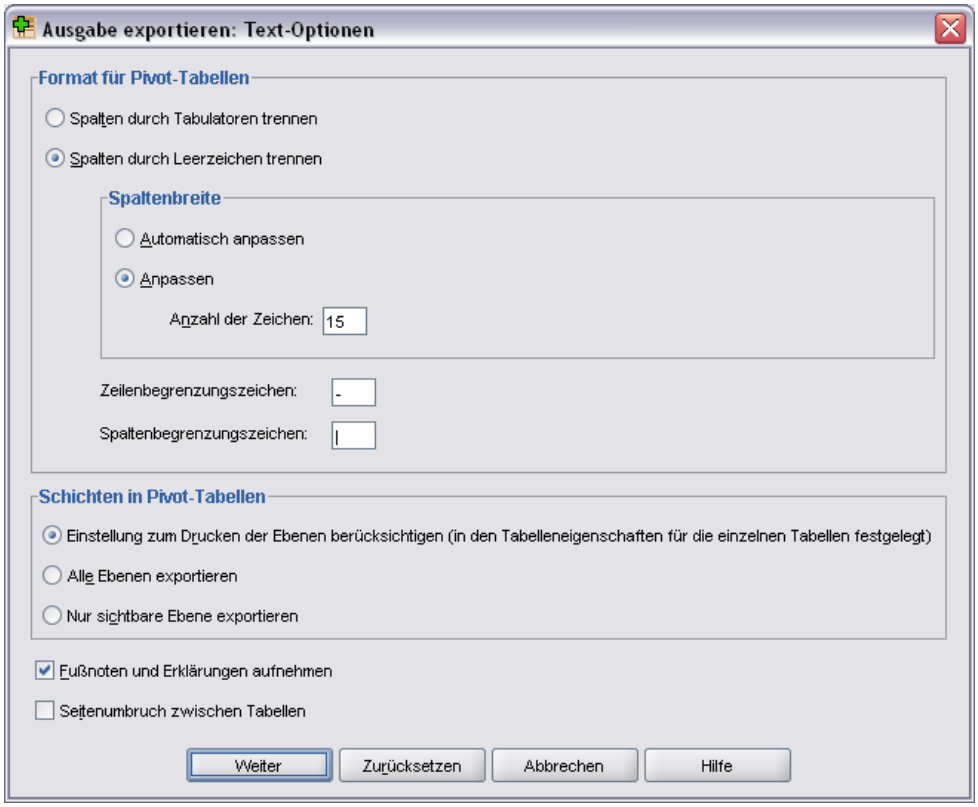
Fußnoten und Erklärungen aufnehmen. Dient zur Festlegung der Aufnahme bzw. des Ausschlusses aller Fußnoten und Erklärungen von Pivot-Tabellen.

Seitenumbruch zwischen Tabellen. Fügt zwischen den einzelnen Tabellen einen Seitenvorschub/-umbruch ein. Mithilfe dieser Option wird bei mehrschichtigen Pivot-Tabellen ein Seitenumbruch zwischen den einzelnen Schichten eingefügt.

So legen Sie Text-Exportoptionen fest:

- ▶ Wählen Sie Text als Exportformat.
- ▶ Klicken Sie auf Optionen ändern.

Abbildung 10-6
Dialogfeld "Text-Optionen"



Optionen zum Exportieren von Diagrammen

Bei HTML- und Textdokumenten und nur zum Exportieren von Diagrammen können Sie das Grafikformat auswählen und für jedes Grafikformat verschiedene optionale Einstellungen festlegen.

So wählen Sie das Grafikformat und die Optionen für exportierte Diagramme aus:

- Wählen Sie HTML, Text oder Ohne (nur Grafiken) als Dokumenttyp aus.
- Wählen Sie das Grafikdateiformat aus der Dropdown-Liste aus.
- Klicken Sie auf Optionen ändern, um die Optionen für das ausgewählte Grafikdateiformat zu ändern.

Exportoptionen für JPEG-Diagramme

- **Bildgröße.** Prozentsatz der ursprünglichen Diagrammgröße, bis zu 200 Prozent.
- **In Graustufen umwandeln.** Wandelt die verschiedenen Farben in Grauschattierungen um.

Exportoptionen für BMP-Diagramme

- **Bildgröße.** Prozentsatz der ursprünglichen Diagrammgröße, bis zu 200 Prozent.
- **Bild zur Verringerung der Dateigröße komprimieren.** Ein verlustfreies Komprimierungsverfahren, das kleinere Dateien ohne Beeinträchtigung der Bildqualität erstellt.

Exportoptionen für PNG-Diagramme

Bildgröße. Prozentsatz der ursprünglichen Diagrammgröße, bis zu 200 Prozent.

Farbtiefe. Bestimmt die Anzahl der Farben im exportierten Diagramm. Ein Diagramm, das mit einer beliebigen Farbtiefe gespeichert wird, verfügt über eine Mindestzahl tatsächlich verwendeter Farben und eine Höchstzahl zulässiger Farben in dieser Farbtiefe. Wenn das Diagramm beispielsweise die drei Farben Rot, Weiß und Schwarz enthält, jedoch als Bild mit 16 Farben gespeichert wird, verbleibt das Diagramm im dreifarbigen Modus.

- Wenn die Anzahl der Farben im Diagramm die Anzahl der Farben für diese Tiefe übersteigt, werden die Farben zur Replizierung der Diagrammfarben gemischt.
- Aktuelle Bildschirmtiefe ist die Anzahl der Farben, die gegenwärtig auf Ihrem Monitor dargestellt werden.

Exportoptionen für EMF- und TIFF-Diagramme

Bildgröße. Prozentsatz der ursprünglichen Diagrammgröße, bis zu 200 Prozent.

Anmerkung: Das Format EMF (Enhanced Metafile) ist nur unter Windows-Betriebssystemen verfügbar.

Exportoptionen für EPS-Diagramme

Bildgröße. Sie können die Größe als Prozentsatz der ursprünglichen Bildgröße (bis zu 200 Prozent) angeben oder Sie können eine Bildbreite in Pixel angeben (dabei richtet sich die Höhe nach dem Wert für die Breite und dem Seitenverhältnis). Das exportierte Bild ist immer proportional zum ursprünglichen Bild.

TIFF-Vorschaubild einschließen. Speichert eine Vorschau mit dem EPS-Bild im TIFF-Format zur Anzeige in Anwendungen, bei denen keine EPS-Bilder auf dem Bildschirm dargestellt werden können.

Schriftarten. Steuert die Behandlung von Schriftarten in EPS-Bildern.

- **Schriftartreferenzen verwenden.** Wenn die im Diagramm verwendeten Schriftarten auf dem Ausgabegerät zur Verfügung stehen, werden sie verwendet. Anderenfalls verwendet das Ausgabegerät andere Schriftarten.
- **Schriftarten durch Kurven ersetzen.** Wandelt Schriftarten in PostScript-Kurvendaten um. Der Text selbst kann dann nicht mehr bearbeitet werden. Diese Option ist sinnvoll, wenn die im Diagramm verwendeten Schriftarten auf dem Ausgabegerät nicht zur Verfügung stehen.

Ausdrucken von Viewer-Dokumenten

Es stehen zwei Optionen zum Drucken des Inhalts des Viewer-Fensters zur Verfügung:

Alle angezeigten Ausgaben. Hiermit werden nur die gegenwärtig im Inhaltsfenster angezeigten Objekte gedruckt. Ausgeblendete Objekte werden nicht gedruckt. (Ausgeblendete Objekte sind Objekte, die im Gliederungsfenster mit einem geschlossenen Buch dargestellt werden oder in reduzierten Gliederungsschichten verborgen sind.)

Auswahl. Hiermit werden nur die gegenwärtig im Gliederungs- und/oder Inhaltsfenster ausgewählten Objekte gedruckt.

So drucken Sie Ausgaben und Diagramme:

- ▶ Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

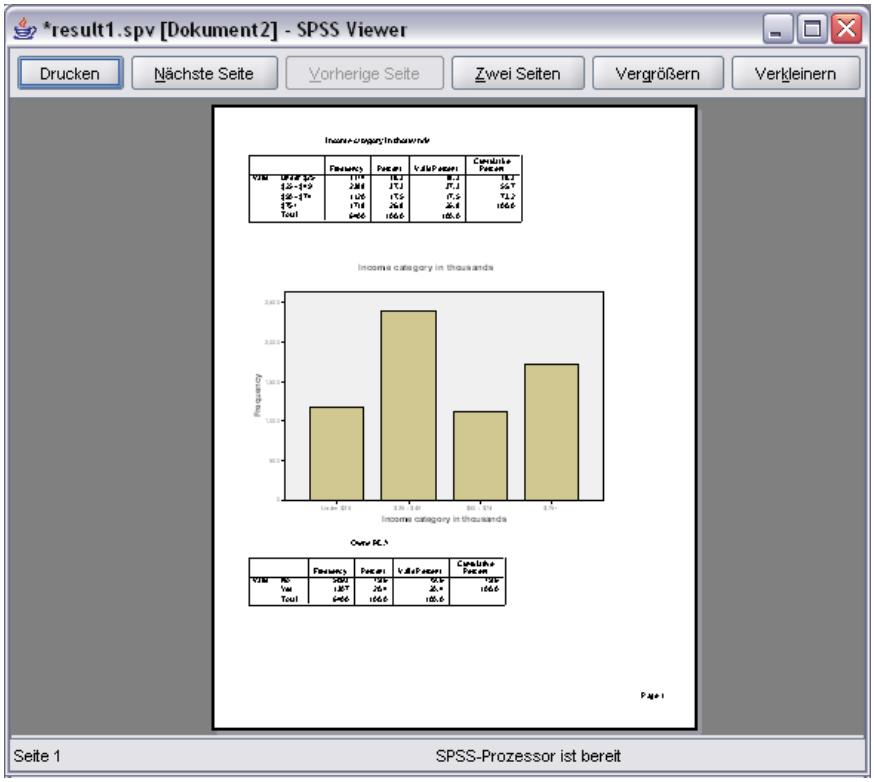
  Datei
  Drucken...
- ▶ Wählen Sie die gewünschten Druckereinstellungen.
- ▶ Klicken Sie zum Drucken auf OK.

Seitenansicht

Mit “Seitenansicht” erhalten Sie für jede Seite eine Vorschau des Ausdrucks von Viewer-Dokumenten. Bevor Sie Viewer-Dokumente drucken, sollten Sie diese in der Seitenansicht überprüfen. In der Seitenansicht werden Objekte angezeigt, die möglicherweise nicht sichtbar sind, wenn Sie nur das Inhaltsfenster des Viewers betrachten, beispielsweise:

- Seitenumbrüche,
- verborgene Schichten von Pivot-Tabellen,
- Umbrüche in breiten Tabellen,
- auf jeder Seite zu druckende Kopf- und Fußzeilen.

Abbildung 10-7
Seitenansicht



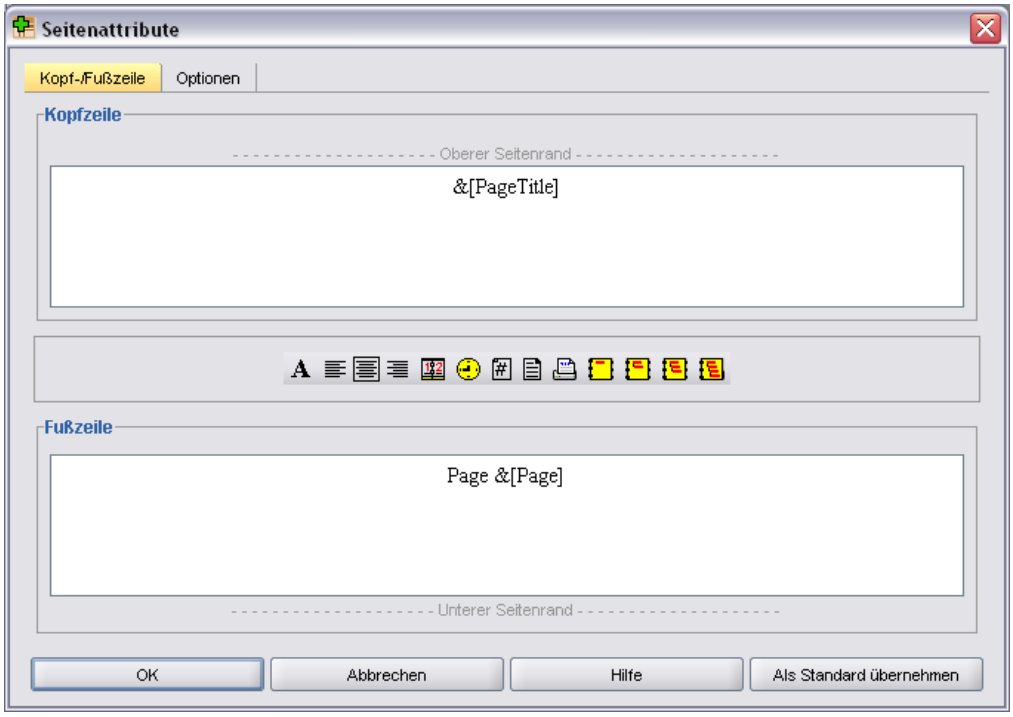
Wenn im Viewer gegenwärtig eine Ausgabe ausgewählt ist, wird in der Seitenansicht nur diese Ausgabe angezeigt. Wenn Sie eine Vorschau für alle Ausgaben sehen möchten, darf im Viewer kein Objekt ausgewählt sein.

Seitenattribute: Kopf-/Fußzeile

Kopf- und Fußzeilen sind die Informationen, die am oberen und unteren Rand jeder Seite ausgedruckt werden. Sie können beliebigen Text als Kopf- und Fußzeile eingeben. Außerdem können Sie die Symbolleiste in der Mitte des Dialogfelds verwenden, wenn Sie folgendes einfügen möchten:

- Datum und Uhrzeit,
- Seitennummern,
- Dateiname aus dem Viewer,
- Beschriftungen der Gliederungsüberschriften,
- Titel und Untertitel.

Abbildung 10-8
Dialogfeld "Seitenattribute", Registerkarte "Kopf-/Fußzeile"



- Als Standard übernehmen verwendet die hier angegebenen Einstellungen als Standardeinstellungen für neue Viewer-Dokumente.
- Mit den Beschriftungen der Gliederungsüberschriften werden die Gliederungsüberschriften der ersten, zweiten, dritten und/oder vierten Stufe für das erste Objekt auf jeder Seite angezeigt.
- Mit den Symbolen für die Seitentitel und die Untertitel werden die aktuellen Titel und Untertitel der Seite gedruckt. Diese Dokumente können mit der Option "Neuer Seitentitel" im Menü "Einfügen" des Viewers erstellt werden oder mit den Befehlen `TITLE` und `SUBTITLE`. Wenn Sie keine Seitentitel oder -untertitel angegeben haben, wird diese Einstellung ignoriert.

  Anmerkung: Die Merkmale der Schriftart für neue Seitentitel und -untertitel werden auf der Registerkarte "Viewer" des Dialogfelds "Optionen" festgelegt (Menü "Bearbeiten", Befehl "Optionen"). Bei bestehenden Seitentiteln und -untertiteln können die Merkmale der Schriftart durch Bearbeiten der Titel direkt im Viewer geändert werden.

Um eine Vorschau des Erscheinungsbilds von Kopf- und Fußzeilen auf der gedruckten Seite zu erhalten, wählen Sie im Menü "Datei" die Option "Seitenansicht" aus.

So fügen Sie Kopf- und Fußzeilen ein:

▶ Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Datei
  Seitenattribute...

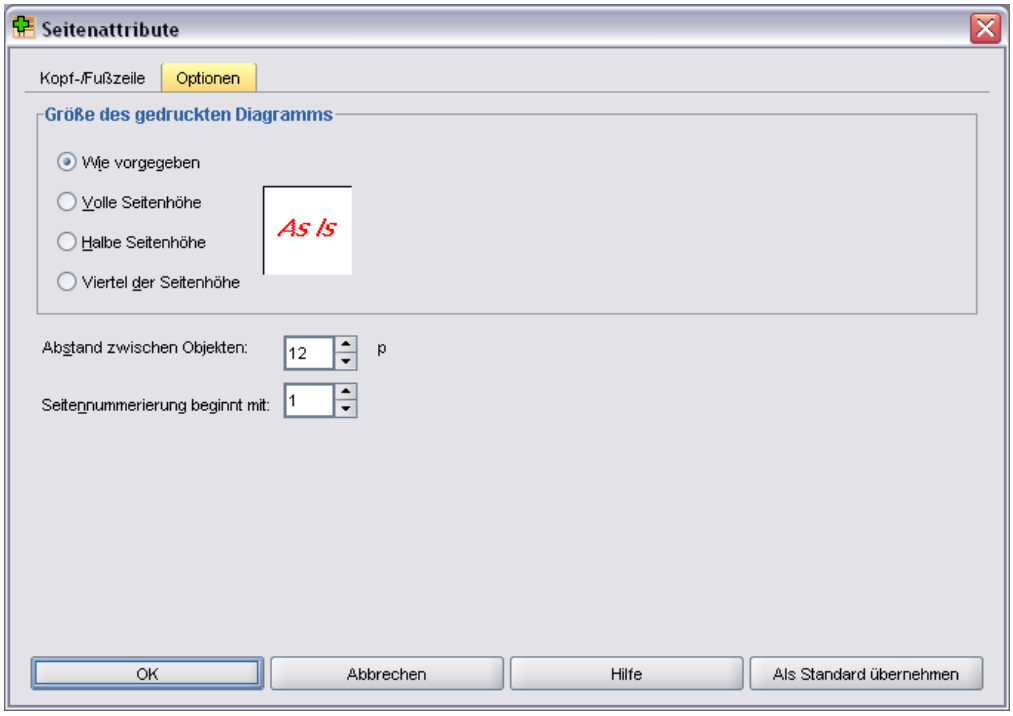
- Klicken Sie auf die Registerkarte Kopf-/Fußzeile.
- Geben Sie die gewünschte Kopf- und/oder Fußzeile ein, die auf jeder Seite angezeigt werden soll.

Seitenattribute: Optionen

In diesem Dialogfeld werden die Größe der gedruckten Diagramme, der Abstand zwischen gedruckten Ausgabeobjekten und die Seitennumerierung eingestellt.

- **Größe des gedruckten Diagramms.** Hier wird die Größe des gedruckten Diagramms im Verhältnis zur definierten Seitengröße festgelegt. Die Größe des gedruckten Diagramms hat keine Auswirkungen auf das Seitenverhältnis (Verhältnis Breite zu Höhe) des Diagramms. Die Gesamtgröße eines gedruckten Diagramms wird von dessen Höhe und Breite bestimmt. Wenn die äußeren Grenzen eines Diagramms den linken und rechten Rand der Seite berühren, kann das Diagramm in der Höhe nicht weiter vergrößert werden.
- **Abstand zwischen Objekten.** Hier wird der Abstand zwischen gedruckten Objekten festgelegt. Jede Pivot-Tabelle, jedes Diagramm und jedes Textobjekt ist ein separates Objekt. Diese Einstellung wirkt sich nicht auf die Anzeige von Objekten im Viewer aus.
- **Seitennumerierung beginnen mit.** Die Seiten werden fortlaufend ab der angegebenen Nummer numeriert.
- **Als Standard übernehmen.** Diese Option verwendet die hier angegebenen Einstellungen als Standardeinstellungen für neue Viewer-Dokumente.

Abbildung 10-9
Dialogfeld "Seitenattribute", Registerkarte "Optionen"



So ändern Sie die Größe des gedruckten Diagramms, die Seitennumerierung und den Abstand zwischen gedruckten Objekten:

- ▶ Wechseln Sie in das Viewer-Fenster. (Klicken Sie auf eine beliebige Stelle im Fenster.)
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Datei
  Seitenattribute...
- ▶ Klicken Sie auf die Registerkarte Optionen.
- ▶ Ändern Sie die Einstellungen und klicken Sie auf OK.

Speichern der Ausgabe

Der Inhalt des Viewers kann in einem Viewer-Dokument gespeichert werden. Das gespeicherte Dokument enthält beide Bereiche des Viewer-Fensters (Gliederung und Inhalt).

So speichern Sie ein Viewer-Dokument:

- ▶ Wählen Sie die folgenden Befehle aus den Menüs des Viewer-Fensters aus:
  Datei
  Speichern
- ▶ Geben Sie einen Namen für das Dokument ein und klicken Sie anschließend auf Speichern.

  Verwenden Sie zum Speichern von Ergebnissen in externen Formaten (zum Beispiel HTML-Format oder Text-Format) den Befehl “Exportieren” im Menü “Datei”.

Kapitel
11

Pivot-Tabellen

Viele Ergebnisse werden in Tabellen dargestellt, die interaktiv pivotiert werden können. Das heißt, Sie können die Zeilen, Spalten und Schichten neu anordnen.

Bearbeiten von Pivot-Tabellen

Für die Bearbeitung von Pivot-Tabellen stehen folgende Optionen zur Verfügung:

- Transponieren von Zeilen und Spalten,
- Verschieben von Zeilen und Spalten,
- Erstellen von mehrdimensionalen Schichten,
- Anlegen und Aufheben von Gruppierungen für Zeilen und Spalten,
- Anzeigen und ausblenden von Zeilen, Spalten und anderen Informationen,
- Drehen von Zeilen- und Spaltenbeschriftungen,
- Anzeigen von Definitionen für Terme.

Aktivieren von Pivot-Tabellen

Bevor Sie eine Pivot-Tabelle bearbeiten oder ändern können, müssen Sie sie zunächst **aktivieren**. So aktivieren Sie eine Tabelle:

▶ Doppelklicken Sie auf die Tabelle.

oder

▶ Klicken Sie mit der rechten Maustaste auf die Tabelle und wählen Sie im Kontextmenü die Option Inhalt bearbeiten aus.

▶ Wählen Sie im Untermenü entweder Im Viewer oder In separatem Fenster aus.

- In der Standardeinstellung werden bei der Aktivierung der Tabelle durch Doppelklicken alle Tabellen (bis auf extrem große Tabellen) im Viewer-Fenster aktiviert. Für weitere Informationen siehe Pivottabellenoptionen in Kapitel 45 auf S. 502.
- Wenn mehrere Pivot-Tabellen gleichzeitig aktiviert sein sollen, müssen Sie die Tabellen in separaten Fenstern aktivieren.

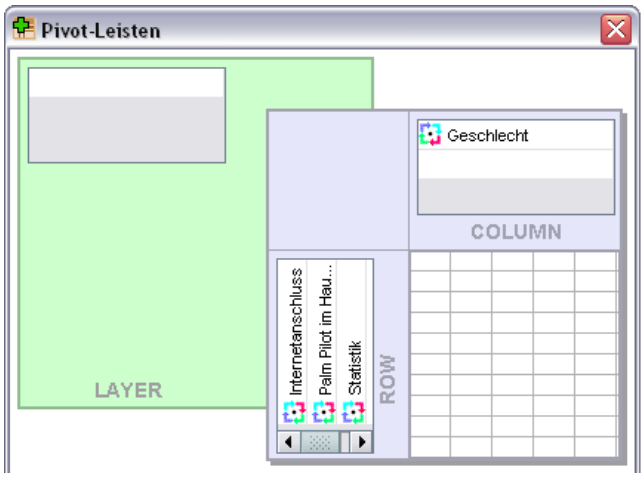
Pivotieren einer Tabelle

▶ Aktivieren Sie die Pivot-Tabelle.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Pivot
  Pivot-Leisten

Abbildung 11-1
Pivot-Leisten



Tabellen weisen drei Dimensionen auf: Zeilen, Spalten und Schichten. Eine Dimension kann mehrere Elemente (oder überhaupt keine) enthalten. Sie können die Anordnung der Tabelle ändern, indem Sie Elemente zwischen den Dimensionen oder innerhalb der Dimensionen verschieben. Sie können die Elemente einfach mittels Ziehen und Ablegen an die gewünschte Position verschieben.

Ändern der Anzeigereihenfolge der Elemente innerhalb einer Dimension

So ändern Sie die Anzeigereihenfolge von Elementen in einer Tabellendimension (Zeile, Spalte bzw. Schicht):

▶ Wenn die Pivot-Leisten noch nicht eingeblendet sind, wählen Sie folgende Befehle aus dem Menü “Pivot-Tabelle” aus:

Pivot
  Pivot-Leisten

▶ Verschieben Sie mittels Ziehen und Ablegen die Elemente innerhalb der Dimension in der Pivot-Leiste.

Verschieben von Zeilen und Spalten innerhalb eines Dimensionselements

▶ Klicken Sie in der Tabelle selbst (nicht in den Pivot-Leisten) auf die Beschriftung der zu verschiebenden Zeile oder Spalte.

▶ Ziehen Sie die Beschriftung an die neue Position.

▶ Wählen Sie im Kontextmenü den Befehl Einfügen vor oder Vertauschen aus.

Anmerkung: Stellen Sie sicher, dass die Option “Zum Kopieren ziehen” im Menü “Bearbeiten” *nicht* aktiviert ist (also nicht mit einem Häkchen gekennzeichnet ist). Wenn die Option “Zum Kopieren ziehen” aktiviert ist, deaktivieren Sie diese.

Vertauschen von Zeilen und Spalten

Wenn Sie einfach nur die Zeilen und Spalten vertauschen möchten, gibt es eine einfache Alternative zur Verwendung der Pivot-Leisten:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Pivot
  Zeilen und Spalten vertauschen

  Damit erzielen Sie dasselbe Ergebnis, wie wenn Sie alle Zeilenelemente in die Spaltendimension und alle Spaltenelemente in die Zeilendimension ziehen würden.

Gruppieren von Zeilen oder Spalten

- Wählen Sie die Beschriftungen für die Zeilen oder Spalten aus, die Sie gruppieren möchten. (Klicken und ziehen Sie mit der Maus oder halten Sie beim Klicken die Umschalt-Taste gedrückt, um mehrere Beschriftungen auszuwählen).
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Gruppieren

  Es wird automatisch eine Gruppenbeschriftung eingefügt. Doppelklicken Sie auf die Gruppenbeschriftung, um den Text zu bearbeiten.

Abbildung 11-2
Zeilen- und Spaltengruppen und Beschriftungen

		Spaltengruppenbeschriftung		
		Weiblich	Männlich	Gesamt
Zeilengruppenbeschriftung	Büro	206	157	363
	Bewachung		27	27
	Management	10	74	84

Anmerkung: Um Zeilen oder Spalten zu einer bestehenden Gruppe hinzuzufügen, müssen Sie zunächst die Gruppierung der Elemente, die sich derzeit in der Gruppe befinden, aufheben. Anschließend können Sie eine neue Gruppe erstellen, die die zusätzlichen Elemente enthält.

Aufheben der Gruppierung von Zeilen oder Spalten

- Klicken Sie auf eine beliebige Stelle in der Gruppenbeschriftung der Zeilen oder Spalten, deren Gruppierung Sie aufheben möchten.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Gruppierung aufheben

  Beim Aufheben der Gruppierung wird automatisch die Gruppenbeschriftung gelöscht.

Drehen von Zeilen- und Spaltenbeschriftungen

Sie können die Beschriftungen für die innersten Spaltenbeschriftungen und die äußersten Zeilenbeschriftungen in einer Tabelle zwischen horizontaler und vertikaler Anzeige drehen.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Format
Innere Spaltenbeschriftungen drehen

oder

Format
Äußere Zeilenbeschriftungen drehen

Abbildung 11-3
Gedrehte Spaltenbeschriftungen

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
kein Schulabschluss	1390	21.7	21.7	21.7
Hauptschulabschluss	1936	30.2	30.2	52.0
Realschulabschluss	1360	21.2	21.2	73.2
Abitur				
Universitätsal				
Gesamt				

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
kein Schulabschluss	1390	21.7	21.7	21.7
Hauptschulabschluss	1936	30.2	30.2	52.0
Realschulabschluss	1360	21.2	21.2	73.2
Abitur	1355	21.2	21.2	94.4
Universitätsabschluss	359	5.6	5.6	100.0
Gesamt	6400	100.0	100.0	

Nur die innersten Spaltenbeschriftungen und die äußersten Zeilenbeschriftungen können gedreht werden.

Arbeiten mit Schichten

Sie können eine separate zweidimensionale Tabelle für jede Kategorie oder Kombination von Kategorien anzeigen lassen. Die Tabelle ist dabei sozusagen in Schichten gestapelt, und nur die oberste Schicht ist sichtbar.

Erstellen und Anzeigen von Schichten

So erstellen Sie Schichten:

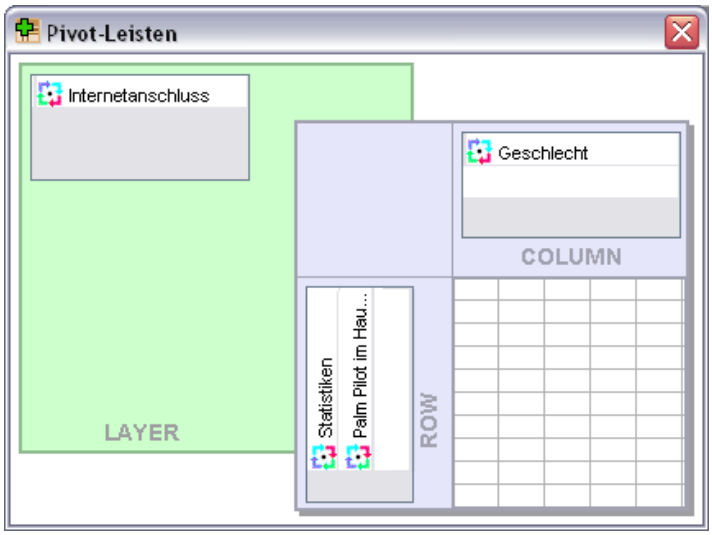
▶ Aktivieren Sie die Pivot-Tabelle.

- Wenn die Pivot-Leisten noch nicht eingeblendet sind, wählen Sie folgende Befehle aus dem Menü “Pivot-Tabelle” aus:

  Pivot
  Pivot-Leisten

- Ziehen Sie ein Element aus der Zeilen- bzw. Spaltendimension in die Schichtdimension.

Abbildung 11-4
Verschieben von Kategorien in Schichten



Durch das Verschieben von Elementen in die Schichtdimension wird eine mehrdimensionale Tabelle erstellt. Es wird jedoch nur ein einziges zweidimensionales Segment angezeigt. Die sichtbare Tabelle ist die Tabelle für die oberste Schicht. Wenn sich beispielsweise eine kategoriale Variable mit den Kategorien “Ja” und “Nein” in der Schichtdimension befindet, enthält die mehrdimensionale Tabelle zwei Schichten: eine für die Kategorie *Ja* und eine für die Kategorie *Nein*.

Abbildung 11-5
Kategorien in separaten Schichten

Internetanschluss: ja

		Anzahl		
		Geschlecht		
		männlich	weiblich	Gesamt
Palm Pilot im Haushalt vorhanden	nein	512	513	1025

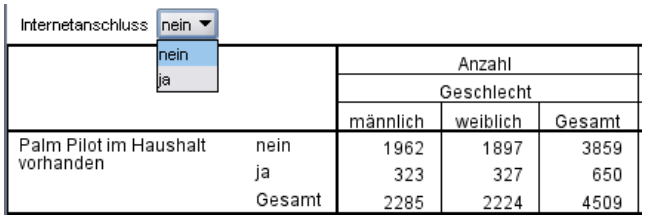
Internetanschluss: nein

		Anzahl		
		Geschlecht		
		männlich	weiblich	Gesamt
Palm Pilot im Haushalt vorhanden	nein	1962	1897	3859
	ja	323	327	650
	Gesamt	2285	2224	4509

Ändern der angezeigten Schicht

▶ Wählen Sie eine Kategorie in der Dropdown-Liste der Schichten aus (in der Pivot-Tabelle selbst, nicht in der Pivot-Leiste).

Abbildung 11-6
Auswählen von Schichten aus Dropdown-Listen



		Anzahl		
		Geschlecht		
		männlich	weiblich	Gesamt
Palm Pilot im Haushalt vorhanden	nein	1962	1897	3859
	ja	323	327	650
	Gesamt	2285	2224	4509

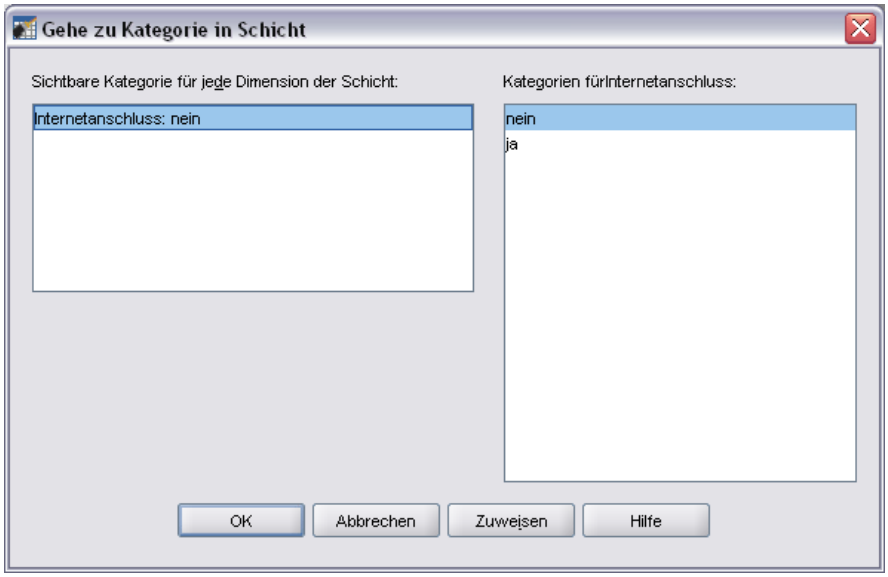
Gehe zu Kategorie in Schicht

Mit “Gehe zu Kategorie in Schicht” können Sie zu einer anderen Schicht in einer Pivot-Tabelle wechseln. Dieses Dialogfeld ist besonders dann nützlich, wenn viele Schichten vorhanden sind oder die ausgewählte Schicht viele Kategorien aufweist.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Pivot
Gehe zu Schicht…

Abbildung 11-7
Dialogfeld “Gehe zu Kategorie in Schicht”



▶ Wählen Sie aus der Liste “Sichtbare Kategorie” eine Schichtendimension aus. In der Liste “Kategorien für Schichten” werden alle Kategorien für die ausgewählte Dimension angezeigt.

▶ Wählen Sie aus der Liste “Kategorien” die gewünschte Kategorie aus und klicken Sie dann auf OK bzw. Zuweisen.

Ein- und Ausblenden von Elementen

Viele Zellentypen können ausgeblendet werden:

- Dimensionsbeschriftungen,
- Kategorien, einschließlich der Zelle mit der Beschriftung und der Datenzellen in einer Zeile oder Spalte,
- Kategoriebeschriftungen (wobei die Datenzellen nicht ausgeblendet werden),
- Fußnoten, Titel und Erklärungen.

Ausblenden von Zeilen und Spalten in einer Tabelle

▶ Klicken Sie mit der rechten Maustaste auf die Kategorienbeschriftung der auszublendenden Zeile oder Spalte.

▶ Wählen Sie folgende Menübefehle aus dem Kontextmenü aus:

Auswählen
  Datenzellen und Beschriftung

▶ Klicken Sie mit der rechten Maustaste erneut auf die Kategorienbeschriftung und wählen Sie im Kontextmenü die Option Kategorie ausblenden aus.

oder

▶ Wählen Sie im Menü "Ansicht" die Option Ausblenden.

Anzeigen ausgeblendeter Zeilen und Spalten in einer Tabelle

▶ Klicken Sie mit der rechten Maustaste auf eine weitere Zeilen- bzw. Spaltenbeschriftung aus derselben Dimension wie die ausgeblendete Zeile oder Spalte.

▶ Wählen Sie folgende Menübefehle aus dem Kontextmenü aus:

Auswählen
  Datenzellen und Beschriftung

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Ansicht
  Alle Kategorien in [Dimensionsname] einblenden

oder

▶ Um alle ausgeblendeten Zeilen und Spalten in einer aktivierten Pivot-Tabelle anzuzeigen, wählen Sie folgende Optionen aus den Menüs aus:

Ansicht
  Alles einblenden

Dadurch werden alle ausgeblendeten Zeilen und Spalten in der Tabelle angezeigt. Wenn Sie aber im Dialogfeld "Tabelleneigenschaften" für diese Tabelle die Option Leere Zeilen und Spalten ausblenden ausgewählt haben, bleiben vollständig leere Zeilen oder Spalten ausgeblendet.

Aus- und Einblenden von Dimensionsbeschriftungen

- ▶ Markieren Sie die Dimensionsbeschriftung oder eine beliebige Kategoriebeschriftung in der Dimension.
- ▶ Wählen Sie im Menü “Ansicht” bzw. im Kontextmenü die Option Beschriftung für Dimension ausblenden bzw. Beschriftung für Dimension einblenden.

Aus- und Einblenden von Tabellentiteln

So blenden Sie einen Titel aus:

- ▶ Wählen Sie den Titel aus.
- ▶ Wählen Sie im Menü “Ansicht” die Option Ausblenden.

So zeigen Sie ausgeblendete Titel an:

- ▶ Wählen Sie im Menü “Ansicht” die Option Alles einblenden.

Tabellenvorlagen

Eine Tabellenvorlage umfaßt ein Set von Eigenschaften, welche die äußere Form einer Tabelle definieren. Sie können eine vordefinierte Tabellenvorlage auswählen oder eine eigene Tabellenvorlage erstellen.

- Vor und nach dem Zuweisen von Tabellenvorlagen können Sie mithilfe der Zelleneigenschaften die Zellenformate einzelner Zellen oder Gruppen von Zellen ändern. Die bearbeiteten Zellenformate bleiben auch nach dem Zuweisen einer neuen Tabellenvorlage erhalten. Für weitere Informationen siehe Zelleneigenschaften auf S. 260.
- Wahlweise können Sie alle Zellen auf die Zellenformate zurücksetzen, die durch die aktuelle Tabellenvorlage definiert sind. Dadurch werden die Formate der Zellen zurückgesetzt, die zuvor bearbeitet wurden. Wenn in der Dateiliste für Tabellenvorlagen Wie angezeigt ausgewählt ist, werden alle bearbeiteten Zellen auf die aktuellen Tabelleneigenschaften zurückgesetzt.

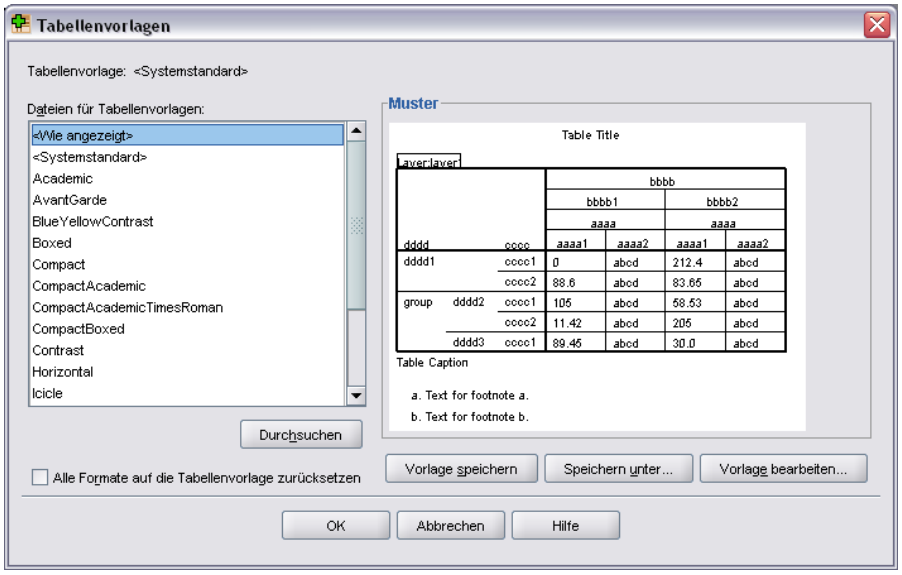
Anmerkung: Tabellenvorlagen, die in früheren Versionen von SPSS erstellt wurden, können nicht in Version 16.0 oder später verwendet werden.

So weisen Sie neue Tabellenvorlagen zu:

- ▶ Aktivieren Sie die Pivot-Tabelle.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Format
  Tabellenvorlagen...

Abbildung 11-8
Dialogfeld "Tabellenvorlagen"



- Wählen Sie eine Tabellenvorlage aus der Liste der Dateien für Tabellenvorlagen aus. Wenn Sie eine Datei in einem anderen Verzeichnis auswählen möchten, klicken Sie auf Durchsuchen.
- Klicken Sie auf OK, um der ausgewählten Pivot-Tabelle die Tabellenvorlage zuzuweisen.

So bearbeiten oder erstellen Sie Tabellenvorlagen:

- Wählen Sie im Dialogfeld "Tabellenvorlagen" aus der Liste der Dateien eine Tabellenvorlage aus.
- Klicken Sie auf Tabellenvorlage bearbeiten.
- Passen Sie die Tabelleneigenschaften an die gewünschten Attribute an und klicken Sie anschließend auf OK.
- Klicken Sie auf Vorlage speichern, um die bearbeitete Tabellenvorlage zu speichern, oder klicken Sie auf Speichern unter, um sie als neue Tabellenvorlage zu speichern.

  Änderungen in einer Tabellenvorlage wirken sich nur auf die ausgewählte Pivot-Tabelle aus. Bearbeitete Tabellenvorlagen werden anderen Tabellen, die diese Tabellenvorlage verwenden, nicht automatisch zugewiesen. Hierfür müssen Sie die betreffenden Tabellen auswählen und die Tabellenvorlage erneut zuweisen.

Tabelleneigenschaften

Unter "Tabelleneigenschaften" können Sie die allgemeinen Eigenschaften einer Tabelle festlegen, Zelleneigenschaften für verschiedene Teile einer Tabelle bestimmen und diese Eigenschaften als Tabellenvorlage speichern. Sie verfügen über folgende Möglichkeiten:

- Festlegen allgemeiner Eigenschaften, beispielsweise das Ausblenden leerer Zeilen und Spalten und das Anpassen der Druckeigenschaften,

- Festlegen des Formats und der Position von Fußnotenzeichen,
- Festlegen spezieller Formate für Zellen im Datenbereich, für Zeilen- und Spaltenbeschriftungen und für andere Bereiche der Tabelle,
- Festlegen der Breite und Farbe für die Rahmenlinien der einzelnen Tabellenbereiche.

So ändern Sie die Eigenschaften von Pivot-Tabellen:

- ▶ Aktivieren Sie die Pivot-Tabelle.
- ▶ Wählen Sie die folgenden Befehle aus den Menüs aus:
  Format
  Tabelleneigenschaften
- ▶ Wählen Sie eine Registerkarte aus (Allgemein, Fußnoten, Zellenformate, Rahmen oder Drucken).
- ▶ Wählen Sie die gewünschten Optionen aus.
- ▶ Klicken Sie auf OK oder Zuweisen.

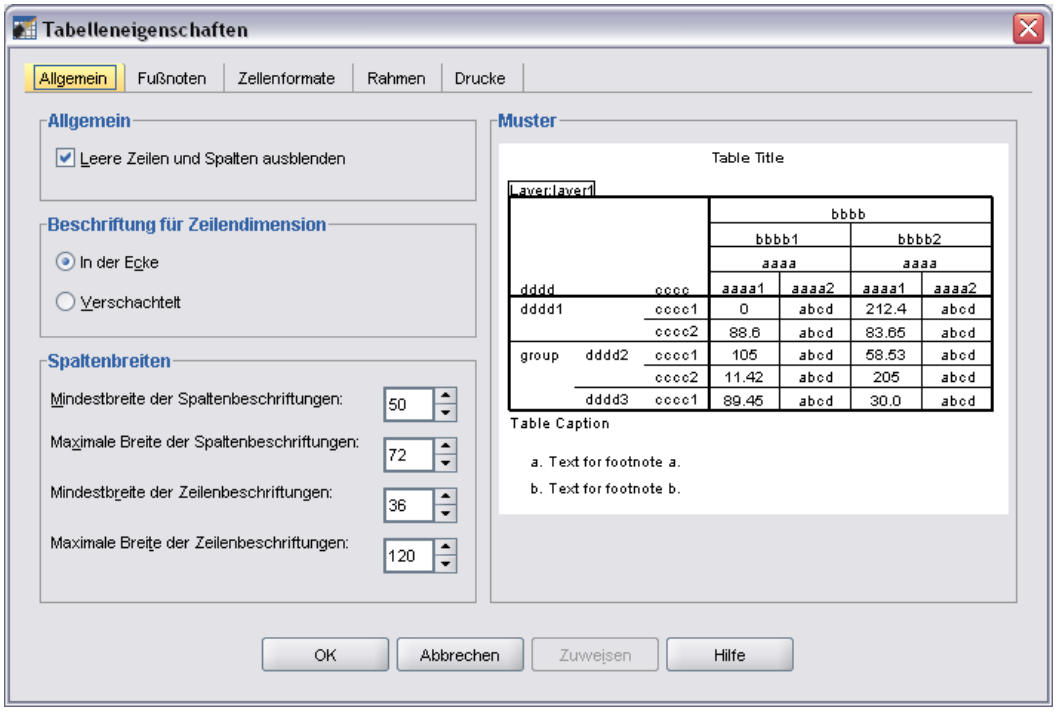
  Die neuen Eigenschaften werden der ausgewählten Pivot-Tabelle zugewiesen. Wenn Sie die neuen Tabelleneigenschaften nicht nur der ausgewählten Tabelle, sondern einer Tabellenvorlage hinzufügen möchten, bearbeiten Sie die Tabellenvorlage. Wählen Sie dazu im Menü “Format” den Befehl “Tabellenvorlage”.

Tabelleneigenschaften: Allgemein

Einige Eigenschaften betreffen die ganze Tabelle. Sie verfügen über folgende Möglichkeiten:

- Ein- oder Ausblenden von leeren Zeilen und Spalten. (Bei einer leeren Zeile oder Spalte steht in keiner der Datenzellen ein Wert.)
- Festlegen der Platzierung der Zeilenbeschriftungen, die sich in der oberen linken Ecke befinden oder verschachtelt sein können.
- Festlegen der maximalen und minimalen Spaltenbreite (angegeben in Punkt).

Abbildung 11-9
Dialogfeld "Tabelleneigenschaften", Registerkarte "Allgemein"



So ändern Sie allgemeine Tabelleneigenschaften:

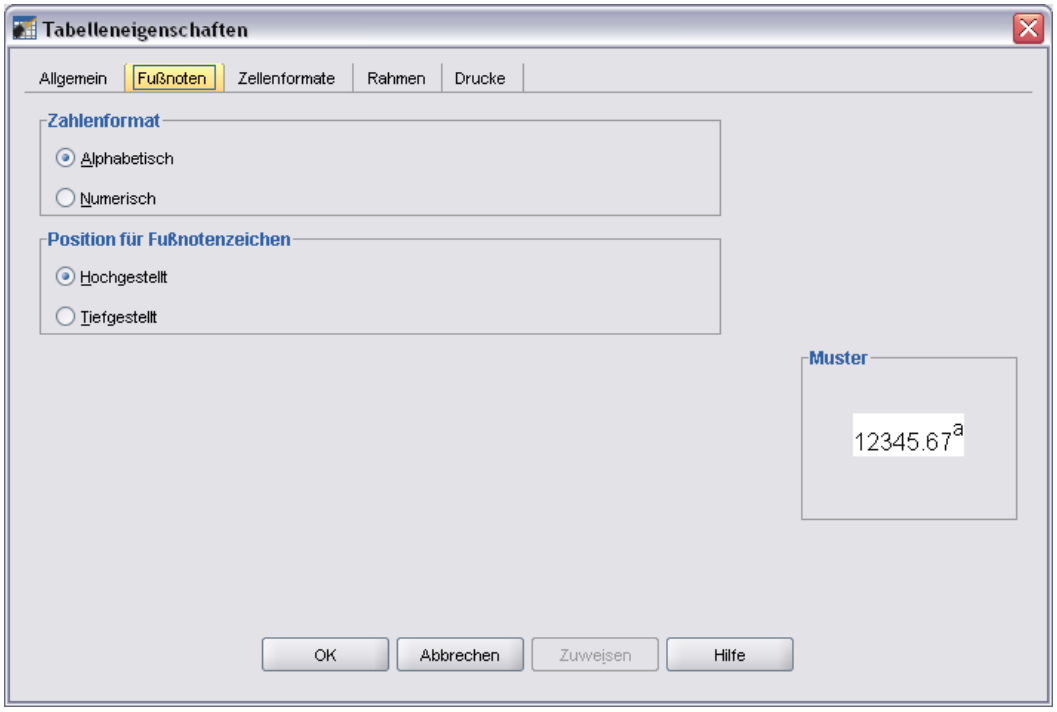
- Klicken Sie auf die Registerkarte Allgemein.
- Wählen Sie die gewünschten Optionen aus.
- Klicken Sie auf OK oder Zuweisen.

Tabelleneigenschaften: Fußnoten

Zu den Eigenschaften von Fußnotenzeichen gehören zum Beispiel Format und Position in Bezug auf den Text.

- Als Format für Fußnotenzeichen können Sie entweder Ziffern (1, 2, 3, ...) oder Buchstaben (a, b, c, ...) festlegen.
- Die Fußnotenzeichen können hochgestellt oder tiefgestellt eingefügt werden.

Abbildung 11-10
Dialogfeld "Tabelleneigenschaften", Registerkarte "Fußnoten"



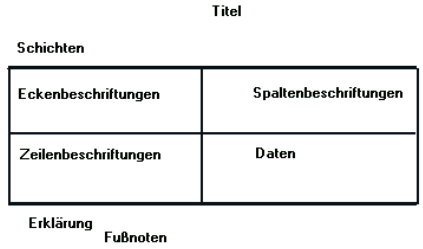
So ändern Sie die Eigenschaften von Fußnotenzeichen:

- Klicken Sie auf die Registerkarte Fußnoten.
- Wählen Sie ein Zahlenformat für die Fußnoten aus.
- Wählen Sie eine Position für Fußnotenzeichen aus.
- Klicken Sie auf OK oder Zuweisen.

Tabelleneigenschaften: Zellenformate

In bezug auf die Formatierung ist eine Tabelle in verschiedene Bereiche aufgeteilt: Titel, Schichten, Eckenbeschriftungen, Zeilenbeschriftungen, Spaltenbeschriftungen, Daten, Erklärungen und Fußnoten. Für jeden Bereich der Tabelle können Sie die entsprechenden Zellenformate ändern. Folgende Zellenformate können bearbeitet werden: Texteigenschaften (z. B. Schriftart, Größe, Farbe und Schnitt), horizontale und vertikale Ausrichtung, Hintergrundfarben sowie die inneren Zellenränder.

Abbildung 11-11
Bereiche einer Tabelle

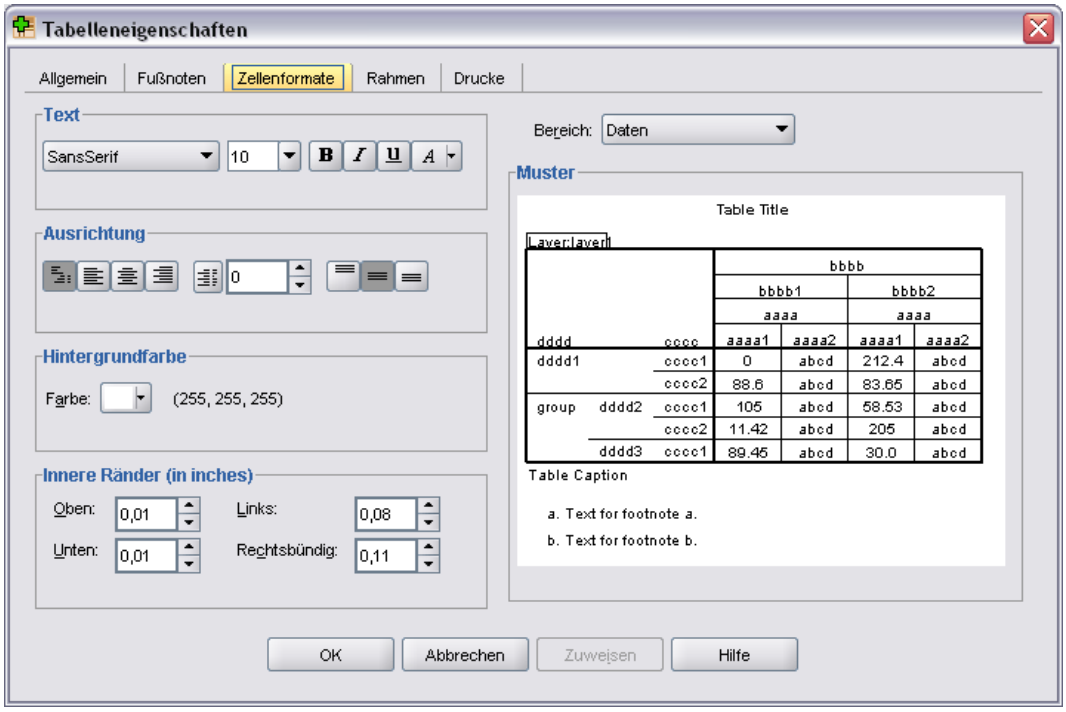


Zellenformate werden immer ganzen Bereichen (Informationskategorien) zugewiesen. Sie sind nicht Eigenschaften einzelner Zellen. Diese Unterscheidung ist besonders beim Pivotieren von Tabellen wichtig.

Beispiel:

- Wenn Sie als Zellenformat für Spaltenbeschriftungen eine fette Schriftart festlegen, werden die Spaltenbeschriftungen fett angezeigt, und zwar unabhängig davon, was gerade in der Spaltendimension angezeigt wird. Wenn Sie ein Element aus der Spaltendimension in eine andere Dimension verschieben, bleibt die für Spaltenbeschriftungen geltende Fettformatierung für dieses Element nicht erhalten.
- Wenn Sie hingegen die Spaltenbeschriftungen fett formatieren, indem Sie die Zellen in einer aktivierten Pivot-Tabelle markieren und auf der Symbolleiste auf die Schaltfläche “Fett” klicken, bleibt der Inhalt dieser Zellen auch beim Verschieben in andere Dimensionen immer fett formatiert. Die Spaltenbeschriftungen behalten diese Formatierung dann nicht für andere Elemente, die in die Spaltendimension verschoben werden.

Abbildung 11-12
Dialogfeld "Tabelleneigenschaften", Registerkarte "Zellenformate"



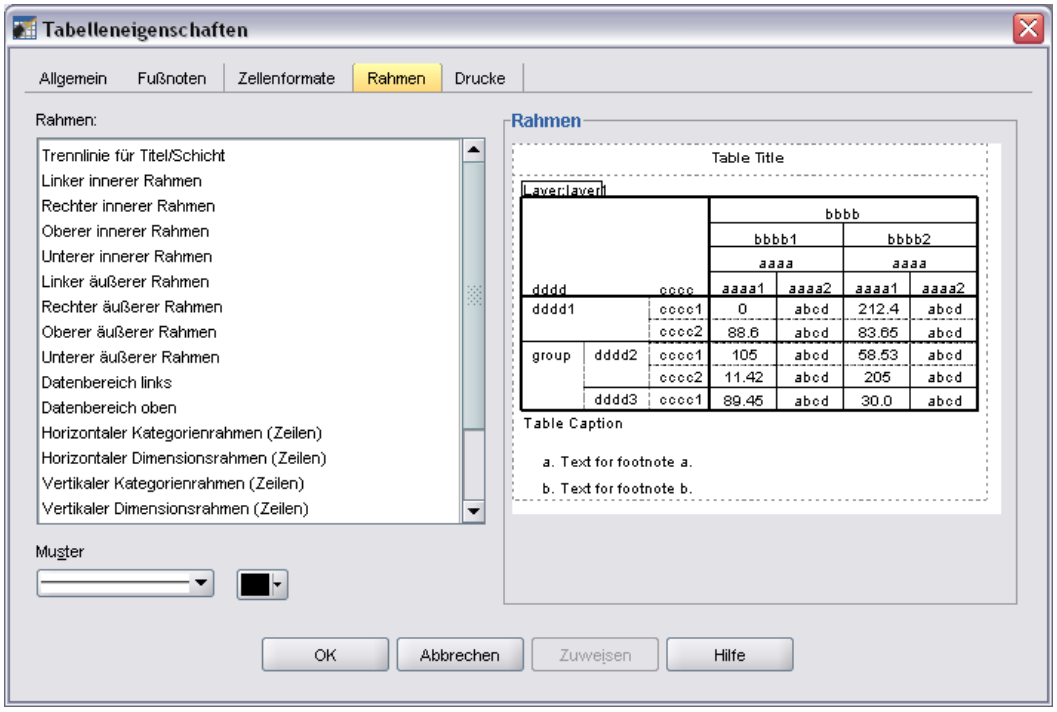
So ändern Sie Zellenformate:

- Klicken Sie auf die Registerkarte Zellenformate.
- Wählen Sie einen Bereich aus der Dropdown-Liste aus oder klicken Sie auf einen Bereich in der Vorschau.
- Wählen Sie die Eigenschaften für diesen Bereich aus. Ihre Auswahl wird in der Vorschau angezeigt.
- Klicken Sie auf OK oder Zuweisen.

Tabelleneigenschaften: Rahmen

Sie können für jeden Rahmen in einer Tabelle einen Linienstil und eine Farbe auswählen. Wenn Sie Kein in der Liste für die Linienstile auswählen, wird an der entsprechenden Position keine Linie gezeichnet.

Abbildung 11-13
Dialogfeld "Tabelleneigenschaften", Registerkarte "Rahmen"



So ändern Sie die Rahmenlinien für Tabellen:

- ▶ Klicken Sie auf die Registerkarte Rahmen.
- ▶ Wählen Sie eine Rahmenposition aus. Klicken Sie hierfür auf den entsprechenden Namen in der Liste oder klicken Sie auf eine Linie in der Vorschau.
- ▶ Wählen Sie einen Linienstil oder wählen Sie die Option Kein aus.
- ▶ Wählen Sie eine Farbe aus.
- ▶ Klicken Sie auf OK oder Zuweisen.

Tabelleneigenschaften: Drucken

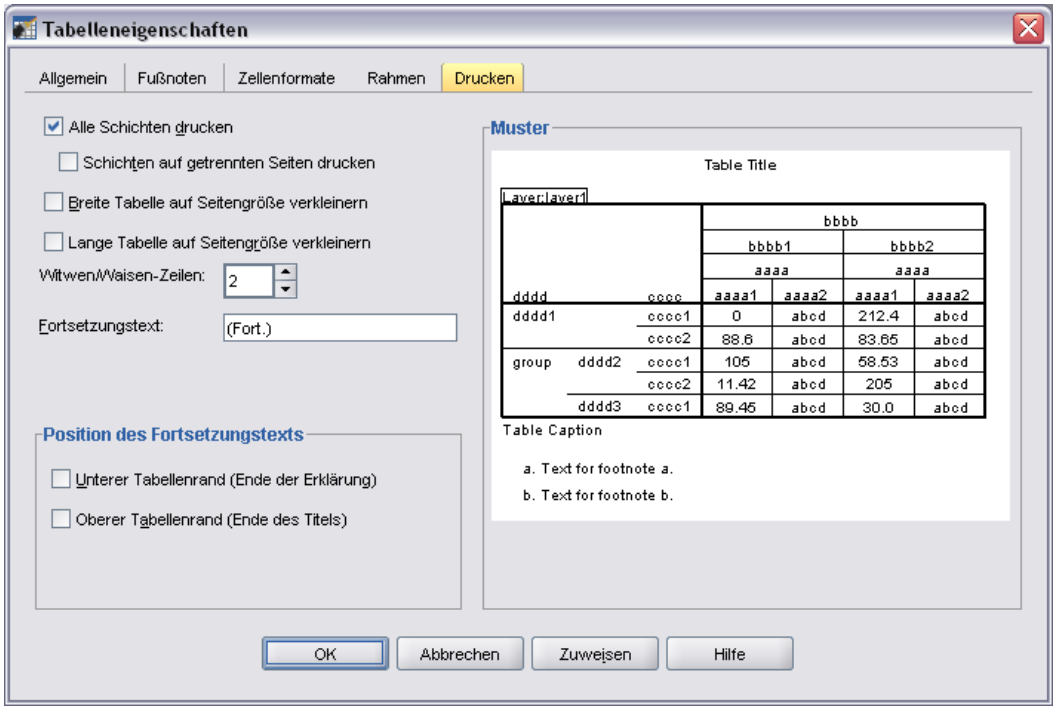
Sie können folgende Eigenschaften für den Ausdruck von Pivot-Tabellen beeinflussen:

- Drucken aller Schichten oder nur der obersten Schicht der Tabelle und Drucken jeder Schicht auf einer getrennten Seite.
- Horizontales oder vertikales Verkleinern einer Tabelle zur Anpassung an die Seitengröße beim Drucken.
- Legen Sie die Einstellungen für Witwen/Waisen-Zeilen fest, in dem Sie die Mindestanzahl von Zeilen und Spalten festlegen, die in einem beliebigen gedruckten Abschnitt einer Tabelle enthalten sind, wenn die Tabelle zu breit und/oder zu lang für die definierte Seitengröße ist.

Anmerkung: Wenn eine Tabelle nicht mehr auf die aktuelle Seite passt, weil sich bereits andere Ausgaben darüber befinden, die Tabelle im Prinzip aber auf eine Seite passen würde, wird sie automatisch auf einer neuen Seite gedruckt. Dies geschieht unabhängig von den Einstellungen für Witwen/Waisen-Zeilen.

- Berücksichtigen von Fortsetzungstexten für Tabellen, die nicht auf eine einzige Seite passen. Fortsetzungstexte können am unteren oder oberen Rand jeder Seite angezeigt werden. Wenn Sie keine dieser Optionen auswählen, werden Fortsetzungstexte nicht angezeigt.

Abbildung 11-14
Dialogfeld "Tabelleneigenschaften", Registerkarte "Drucken"



So wählen Sie Druckeigenschaften für Pivot-Tabellen aus:

- ▶ Klicken Sie auf die Registerkarte Drucken.
- ▶ Wählen Sie die gewünschten Druckoptionen aus.
- ▶ Klicken Sie auf OK oder Zuweisen.

Zelleneigenschaften

Zelleneigenschaften werden ausgewählten Zellen zugewiesen. Sie können Schriftart, Werteformat, Ausrichtung, Ränder und Farben ändern. Zelleneigenschaften setzen Tabelleneigenschaften außer Kraft. Wenn Sie also die Tabelleneigenschaften ändern, bleiben vorher zugewiesene Zelleneigenschaften erhalten.

So ändern Sie Zelleneigenschaften:

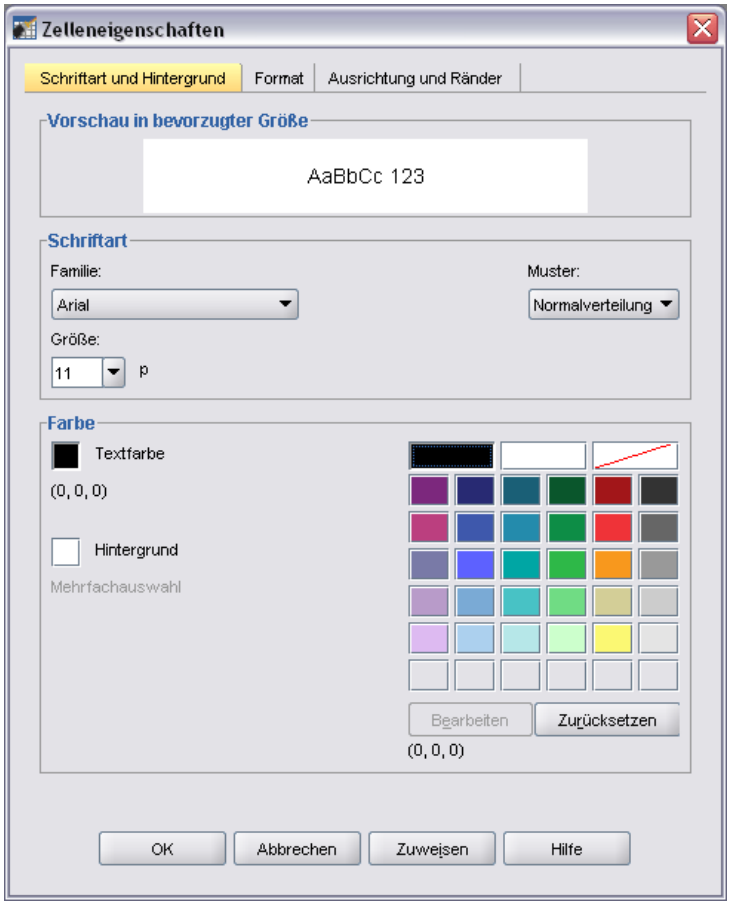
- ▶ Aktivieren Sie eine Tabelle und wählen Sie eine oder mehrere Zellen in der Tabelle aus.

▶ Wählen Sie im Menü “Format” bzw. im Kontextmenü die Option Zelleneigenschaften.

Schriftart und Hintergrund

Mit der Registerkarte “Schriftart und Hintergrund” werden Schriftschnitt und -farbe sowie die Hintergrundfarbe für die ausgewählten Zellen in der Tabelle festgelegt.

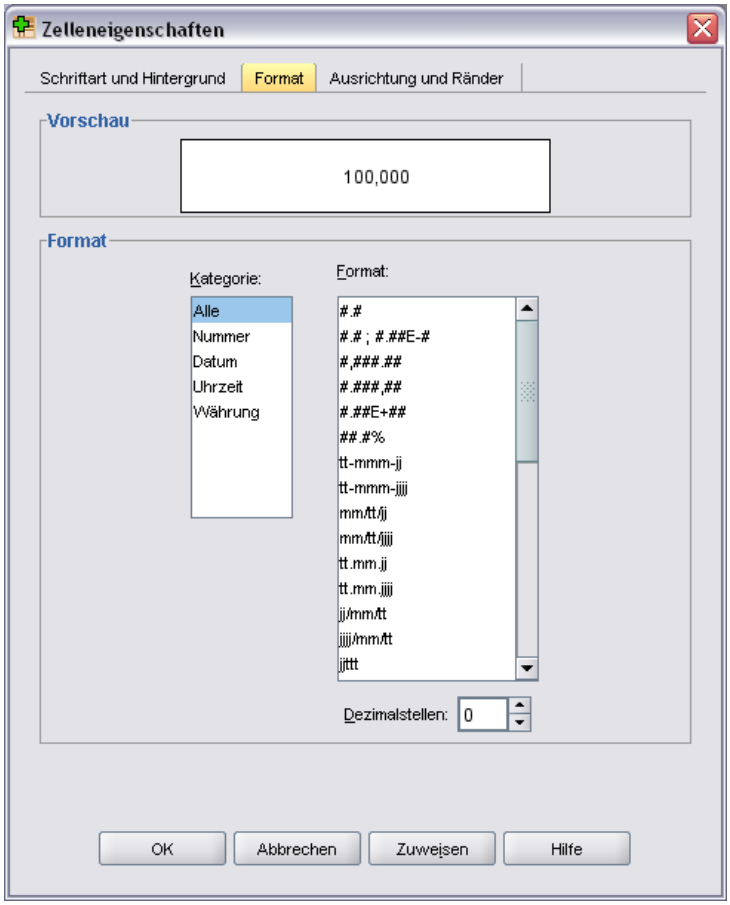
Abbildung 11-15
Dialogfeld “Zelleneigenschaften”, Registerkarte “Schriftart und Hintergrund”



Formatwert

Auf der Registerkarte “Formatwert” werden die Werteformate für die ausgewählten Zellen festgelegt. Sie können die Formate für Zahlen, Datum, Zeit und Währungen bestimmen und die Anzahl der angezeigten Dezimalstellen festlegen.

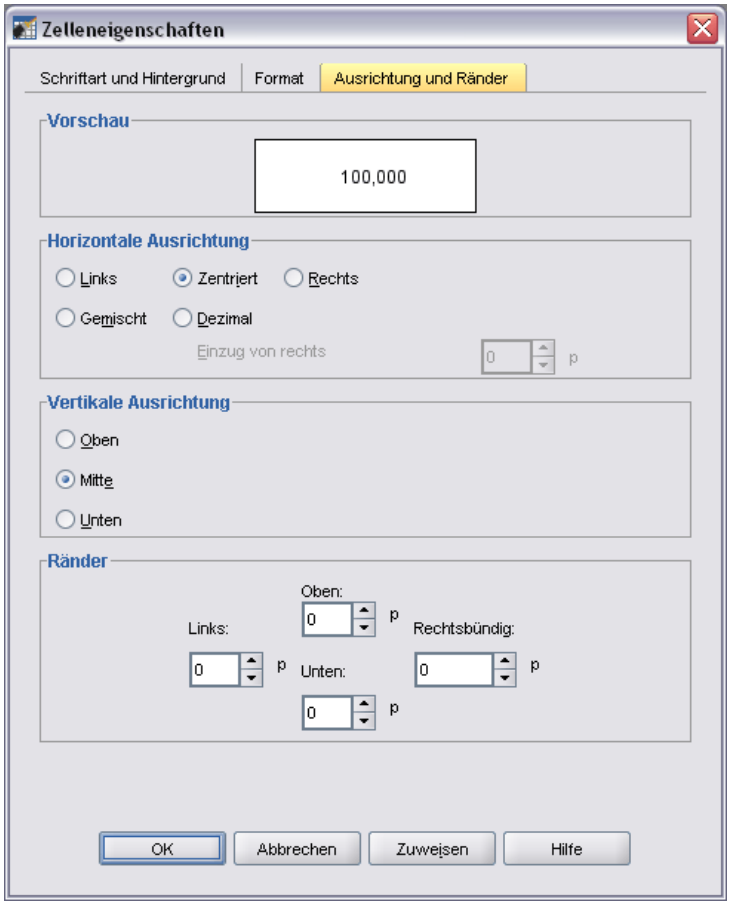
Abbildung 11-16
Dialogfeld "Zelleneigenschaften", Registerkarte "Format"



Ausrichtung und Ränder

Auf der Registerkarte "Ausrichtung und Ränder" können Sie die horizontale und vertikale Ausrichtung von Werten sowie den oberen, unteren, rechten und linken Rand für die ausgewählten Zellen festlegen. Bei einer horizontalen Ausrichtung vom Typ Gemischt werden die Inhalte der einzelnen Zellen gemäß ihrem jeweiligen Typ ausgerichtet. So sind beispielsweise Datumsangaben rechtsbündig und Textwerte linksbündig ausgerichtet.

Abbildung 11-17
Dialogfeld "Zelleneigenschaften", Registerkarte "Ausrichtung und Ränder"



Fußnoten und Erklärungen

Sie können Fußnoten und Erklärungen zu einer Tabelle hinzufügen. Außerdem können Sie Fußnoten bzw. Erklärungen ausblenden, Fußnotenzeichen ändern und Fußnoten neu nummerieren.

Einige Fußnotenattribute richten sich nach den Tabelleneigenschaften. Für weitere Informationen siehe Tabelleneigenschaften: Fußnoten auf S. 255.

Hinzufügen von Fußnoten und Erklärungen

So fügen Sie Tabellen Erklärungen hinzu:

- Wählen Sie im Menü "Einfügen" die Option Erklärung.

Sie können jedem Element einer Tabelle eine Fußnote hinzufügen. So fügen Sie eine Fußnote hinzu:

- Klicken Sie auf einen Titel, eine Zelle oder eine Erklärung in einer aktivierten Pivot-Tabelle.

- Wählen Sie im Menü “Einfügen” die Option Fußnote.

So können Sie eine Erklärung aus- bzw. einblenden:

So blenden Sie eine Erklärung aus:

- Wählen Sie die Erklärung aus.
- Wählen Sie im Menü “Ansicht” die Option Ausblenden.

So zeigen Sie ausgeblendete Erklärungen an:

- Wählen Sie im Menü “Ansicht” die Option Alles einblenden.

So blenden Sie Fußnoten in Tabellen ein und aus:

So blenden Sie eine Fußnote aus:

- Wählen Sie die betreffende Fußnote aus.
- Wählen Sie im Menü “Ansicht” die Option Ausblenden bzw. im Kontextmenü die Option Fußnoten ausblenden.

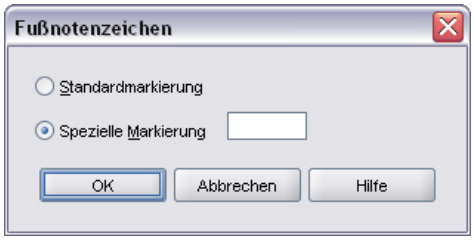
So zeigen Sie ausgeblendete Fußnoten an:

- Wählen Sie im Menü “Ansicht” die Option Alle Fußnoten einblenden.

Fußnotenzeichen

Mit dem Befehl “Fußnotenzeichen” ändern Sie die Zeichen, die zum Kennzeichnen von Fußnoten verwendet werden.

Abbildung 11-18
Dialogfeld “Fußnotenzeichen”



So ändern Sie Fußnotenzeichen:

- Wählen Sie eine Fußnote aus.
- Wählen Sie im Menü “Format” die Option Fußnotenzeichen.
- Geben Sie ein oder zwei Zeichen ein.

Neunummerierung von Fußnoten

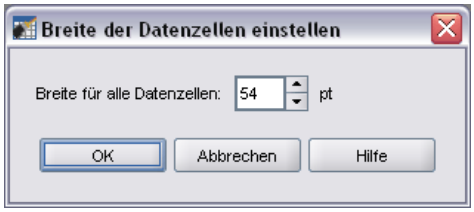
Nach dem Pivotieren einer Tabelle durch Verschieben von Zeilen, Spalten und Schichten ist die Reihenfolge der Fußnoten unter Umständen durcheinander. So numerieren Sie die Fußnoten neu:

- Wählen Sie im Menü “Format” die Option Fußnoten neu nummerieren.

Breite der Datenzellen

Mit “Breite der Datenzellen” können Sie für alle Datenzellen dieselbe Breite festlegen.

Abbildung 11-19
Dialogfeld “Breite der Datenzellen einstellen”



So legen Sie die Breite für alle Datenzellen fest:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Format
  Breite der Datenzellen...

- Geben Sie einen Wert für die Zellenbreite an.

Ändern der Spaltenbreite

- Klicken Sie auf die Spaltenbegrenzung und ziehen Sie daran.

Anzeigen der ausgeblendeten Rahmen in einer Pivot-Tabelle:

Bei Tabellen, bei denen nur wenige Rahmen sichtbar sind, können Sie die ausgeblendeten Rahmen anzeigen lassen. Dadurch werden Vorgänge wie das Ändern von Spaltenbreiten vereinfacht.

- Wählen Sie im Menü “Ansicht” die Option Gitterlinien.

Abbildung 11-20
Gitterlinien für ausgeblendete Rahmen

Internetanschluss				Geschlecht		
				männlich	weiblich	Gesamt
nein	Palm Pilot im Haushalt vorhanden	nein	Anzahl	1962	1897	3859
			Erwartete Anzahl	1955,6	1903,4	3859,0
			Korrigierte Residuen	,5	-,5	
		ja	Anzahl	323	327	650

Internetanschluss				Geschlecht		
				männlich	weiblich	Gesamt
nein	Palm Pilot im Haushalt vorhanden	nein	Anzahl	1962	1897	3859
			Erwartete Anzahl	1955,6	1903,4	3859,0
			Korrigierte Residuen	,5	-,5	
		ja	Anzahl	323	327	650
			Erwartete Anzahl	329,4	320,6	650,0
			Korrigierte Residuen	-,5	,5	
		Gesamt	Anzahl	2285	2224	4509
			Erwartete Anzahl	2285,0	2224,0	4509,0
ja	Palm Pilot im Haushalt vorhanden	nein	Anzahl	512	513	1025
			Erwartete Anzahl	512,5	512,5	1025,0
			Korrigierte Residuen	,0	,1	
		ja	Anzahl	306	305	611
			Erwartete Anzahl	305,5	305,5	611,0
			Korrigierte Residuen	,1	,0	
		Gesamt	Anzahl	818	818	1636
			Erwartete Anzahl	818,0	818,0	1636,0

Auswählen von Zeilen und Spalten in Pivot-Tabellen

In Pivot-Tabellen gibt es einige Beschränkungen beim Auswählen ganzer Zeilen oder Spalten. Die optische Markierung, welche die markierte Zeile oder Spalte kennzeichnet, kann gegebenenfalls nicht zusammenhängende Bereiche der Tabelle umfassen. So wählen Sie eine ganze Zeile bzw. Spalte aus:

- Klicken Sie auf eine Zeilen- oder Spaltenbeschriftung.
- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Auswählen
  Datenzellen und Beschriftung

  oder

- Klicken Sie mit der rechten Maustaste auf die Kategoriebeschriftung der Zeile oder Spalte.
- Wählen Sie folgende Menübefehle aus dem Kontextmenü aus:

  Auswählen
  Datenzellen und Beschriftung

  oder

- Klicken Sie bei gedrückter Strg- und Alt-Taste auf die Zeilen- bzw. Spaltenbeschriftung.

Drucken von Pivot-Tabellen

Das Aussehen gedruckter Pivot-Tabellen wird von mehreren Faktoren bestimmt. Diese Faktoren können Sie durch Ändern der Attribute für Pivot-Tabellen beeinflussen.

- Bei multidimensionalen Pivot-Tabellen (Tabellen mit Schichten) können Sie entweder alle Schichten oder nur die oberste (sichtbare) Schicht ausdrucken. Für weitere Informationen siehe Tabelleneigenschaften: Drucken auf S. 259.
- Sie können zu lange oder zu breite Pivot-Tabellen automatisch an die Seitengröße anpassen. Sie können auch die Positionen von Tabellenumbrüchen und Seitenumbrüchen festlegen. Für weitere Informationen siehe Tabelleneigenschaften: Drucken auf S. 259.
- Bei Tabellen, die zu breit oder zu lang für eine einzelne Seite sind, können Sie die Position der Tabellenumbrüche zwischen den Seiten festlegen.

Verwenden Sie den Befehl “Seitenansicht” im Menü “Datei”, um sich anzeigen zu lassen, wie die Pivot-Tabellen auf der gedruckten Seite aussehen würden.

Festlegen von Tabellenumbrüchen für breite und lange Tabellen

Pivot-Tabellen, die zu breit oder zu lang sind, um innerhalb der definierten Seitengröße gedruckt zu werden, werden automatisch aufgeteilt und in mehreren Abschnitten gedruckt. Bei breiten Tabellen werden mehrere Abschnitte auf derselben Seite gedruckt, wenn genügend Platz zur Verfügung steht. Sie verfügen über folgende Möglichkeiten:

- Festlegen der Zeilen und Spalten, an denen große Tabellen geteilt werden,
- Angeben der Zeilen und Spalten, bei denen Tabellen nicht getrennt werden sollen,
- Anpassen von umfangreichen Tabellen an die definierte Seitengröße.

So legen Sie Zeilen- und Spaltenumbrüche für gedruckte Pivot-Tabellen fest:

▶ Klicken Sie auf die Spaltenbeschriftung links neben der Stelle oder klicken Sie auf die Zeilenbeschriftung über der Stelle, an der Sie den Umbruch einfügen möchten.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Format
  Umbruch hier

So legen Sie Zeilen oder Spalten fest, die nicht getrennt werden sollen:

▶ Wählen Sie die Beschriftungen der Zeilen oder Spalten aus, die nicht getrennt werden sollen. (Sie können mehrere Zeilen- oder Spaltenbeschriftungen auswählen, indem Sie mit gedrückter Maustaste ziehen oder die Umschalttaste gedrückt halten und auf die erste und die letzte auszuwählende Beschriftung klicken.)

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Format
  Zusammenhalten

Erstellen eines Diagramms aus einer Pivot-Tabelle

- Doppelklicken Sie auf die Pivot-Tabelle, um diese zu aktivieren.
- Wählen Sie die Zeilen, Spalten oder Zellen aus, die im Diagramm angezeigt werden sollen.
- Klicken Sie mit der rechten Maustaste auf eine beliebige Stelle im ausgewählten Bereich.
- Wählen Sie im Kontextmenü die Option Diagramm erstellen und wählen Sie einen Diagrammtyp aus.

Kapitel
12

Arbeiten mit der Befehlssyntax

Mit der leistungsfähigen Befehlssprache können Sie viele häufig durchzuführende Aufgaben speichern und automatisieren. Sie bietet außerdem einige Funktionen, die nicht über die Menüs und Dialogfelder zur Verfügung stehen.

Auf die meisten SPSS-Befehle können Sie über die Menüs und Dialogfelder zugreifen. Einige Befehle und Optionen sind aber nur in der SPSS-Befehlssprache verfügbar. Mit der Befehlssprache verfügen Sie außerdem über die Möglichkeit, Jobs in einer Syntaxdatei zu speichern. Sie können eine Analyse dann zu einem späteren Zeitpunkt wiederholen oder diese automatisch über einen Produktionsjob ausführen lassen.

Eine Syntaxdatei ist eine einfache Textdatei, die SPSS-Befehle enthält. Es ist zwar möglich, ein Syntax-Fenster zu öffnen und Befehle einzugeben. Mit den folgenden Funktionen kann SPSS Sie jedoch beim Erstellen einer Syntaxdatei unterstützen:

- Einfügen von Syntaxbefehlen aus Dialogfeldern
- Kopieren der Syntax aus dem Ausgabe-Log
- Kopieren der Syntax aus der Journaldatei

Detaillierte Informationen zur Befehlssyntax sind auf zwei Arten verfügbar: als Bestandteil der umfassenden Hilfesystems und als separates Dokument im PDF-Format im Handbuch *Command Syntax Reference*, das auch über das Menü “Hilfe” verfügbar ist.

Sie können auf kontextsensitive Hilfe für den aktuellen Befehl in einem Syntax-Fenster zugreifen, indem Sie die F1-Taste drücken.

Regeln für die Syntax

Wenn Sie während einer Sitzung Befehle in einem Befehlssyntax-Fenster ausführen, werden diese Befehle im **interaktiven** Modus ausgeführt.

Die folgenden Regeln gelten für die Angaben von Befehlen im interaktiven Modus:

- Jeder Befehl muss auf einer neuen Zeile beginnen. Befehle können in jeder Spalte einer Zeile beginnen und für beliebig viele Zeilen fortgesetzt werden. Die einzige Ausnahme ist der Befehl `END DATA`, der in der ersten Spalte der ersten Zeile nach dem Ende der Daten beginnen muss.
- Jeder Befehl muss mit einem Punkt abgeschlossen werden. Es empfiehlt sich jedoch, den Befehlsabschluss bei `BEGIN DATA` wegzulassen, sodass in der Zeile enthaltene Daten als eine einzige, fortlaufende Angabe behandelt werden.
- Der Befehlsabschluss muss das letzte nichtleere Zeichen in einem Befehl sein.
- Falls kein Punkt als Befehlsabschluss vorhanden ist, wird eine leere Zeile als Befehlsabschluss interpretiert.

Anmerkung: Um die Kompatibilität mit anderen Modi für die Ausführung von Befehlen (einschließlich Befehlsdateien, die mit den Befehlen INSERT oder INCLUDE in einer aktiven Sitzung ausgeführt werden) zu wahren, darf keine Zeile der Befehlssyntax mehr als 256 Zeichen umfassen.

- Die meisten Unterbefehle werden durch Schrägstriche (/) voneinander getrennt. Der Schrägstrich vor dem ersten Unterbefehl ist in der Regel optional.
- Variablennamen müssen vollständig ausgeschrieben werden.
- Text in Apostrophen oder Anführungszeichen muss sich auf einer Zeile befinden.
- Zum Kennzeichnen der Dezimalstellen muss unabhängig von den regionalen oder Gebietsschemaeinstellungen der Punkt (.) verwendet werden.
- Variablenamen, die mit einem Punkt enden, können bei Befehlen, die aus einem Dialogfeld übernommen wurden, Fehler hervorrufen. Sie dürfen solche Variablennamen nicht in Dialogfeldern verwenden und sollten sie auch generell vermeiden.

Bei der SPSS-Befehlssyntax wird nicht zwischen Groß- und Kleinschreibung unterschieden. Für viele Befehle können Abkürzungen aus drei oder vier Zeichen verwendet werden. Sie können beliebig viele Zeilen zur Angabe eines einzelnen Befehls verwenden. An fast jedem Punkt, an dem ein Leerzeichen zulässig ist, können Sie beliebig viele Leerzeilen oder Zeilenumbrüche einfügen, beispielsweise bei Schrägstrichen, runden Klammern, arithmetischen Operatoren oder zwischen Variablennamen. Beispiel:

```
FREQUENCIES
  VARIABLES=TÄTIG GESCHL
  /PERCENTILES=25 50 75
  /BARCHART.
```

und

```
freq var=tätig geschl /percent=25 50 75 /bar.
```

Mit beiden Formen wird das gleiche Ergebnis erzielt.

INCLUDE-Dateien

Befehlsdateien, die mit dem Befehl INCLUDE ausgeführt werden, unterliegen den Syntaxregeln für den Stapelverarbeitungsmodus.

Die folgenden Regeln gelten für die Angabe von Befehlen im Stapelverarbeitungs- und Produktionsmodus:

- Alle Befehle in der Befehlsdatei müssen in Spalte 1 beginnen. Sie können in der ersten Spalte Plus- (+) oder Minuszeichen (–) verwenden, wenn Sie Befehle einrücken möchten, um die Befehlsdatei lesbarer zu machen.
- Wenn für einen Befehl mehrere Zeilen verwendet werden, muss Spalte 1 jeder folgenden Zeile leer sein.
- Befehlsabschlüsse sind optional.
- Zeilen dürfen 256 Byte nicht überschreiten. Jedes zusätzliche Zeichen wird abgeschnitten.

Falls noch keine Befehlsdateien vorliegen, in denen der Befehl `INCLUDE` bereits verwendet wird, sollten Sie stattdessen eher den Befehl `INSERT` verwenden, weil hiermit Befehlsdateien berücksichtigt werden, die beiden Regelsätzen entsprechen. Wenn Sie die Befehlssyntax mit der Schaltfläche "Einfügen" aus einem Dialogfeld in das Syntax-Fenster übernehmen, ist das Format für alle Betriebsmodi geeignet. Weitere Informationen finden Sie in der *Befehlssyntax-Referenz* (im PDF-Format über das Menü "Hilfe").

Übernehmen der Befehlssyntax aus Dialogfeldern

Am einfachsten erstellen Sie eine Befehlssyntax-Datei, indem Sie die entsprechenden Optionen in einem SPSS-Dialogfeld auswählen und die Syntax für diese Auswahl in ein Syntax-Fenster übernehmen. Sie können eine Job-Datei erstellen, indem Sie die Syntax einer längeren Analyse Schritt für Schritt übernehmen. Mit dieser Datei können Sie die Analyse zu einem späteren Zeitpunkt wiederholen oder einen Job mit dem SPSS-Produktionsmodus ausführen lassen.

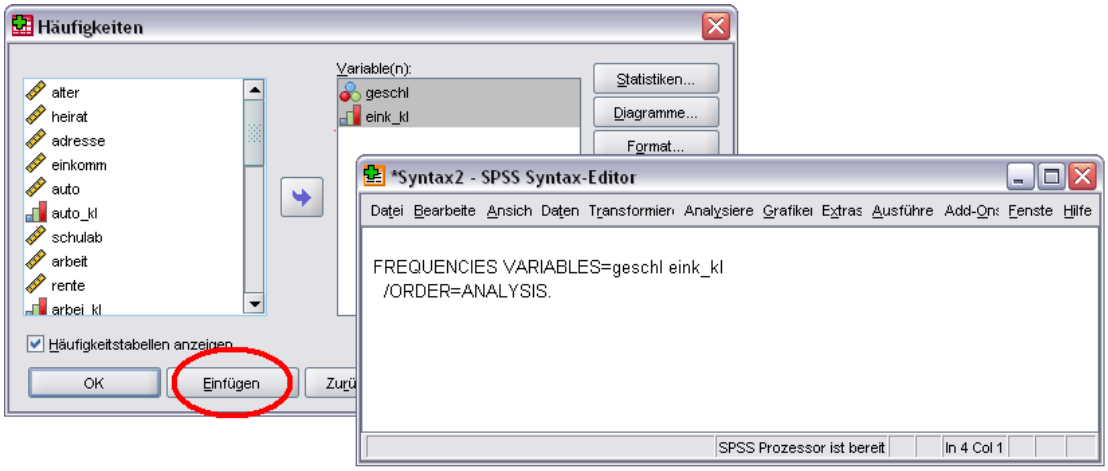
In einem Syntax-Fenster können Sie die übernommene Syntax ausführen, bearbeiten und in einer Syntaxdatei speichern.

So übernehmen Sie die Befehlssyntax aus Dialogfeldern:

- Öffnen Sie das Dialogfeld, und treffen Sie die gewünschte Auswahl.
- Klicken Sie auf Einfügen.

Die Befehlssyntax wird in das Haupt-Syntax-Fenster eingefügt. Falls kein Syntax-Fenster geöffnet ist, öffnet SPSS eines neues Syntax-Fenster und fügt die Syntax dort ein.

Abbildung 12-1
Aus einem Dialogfeld eingefügte Befehlssyntax



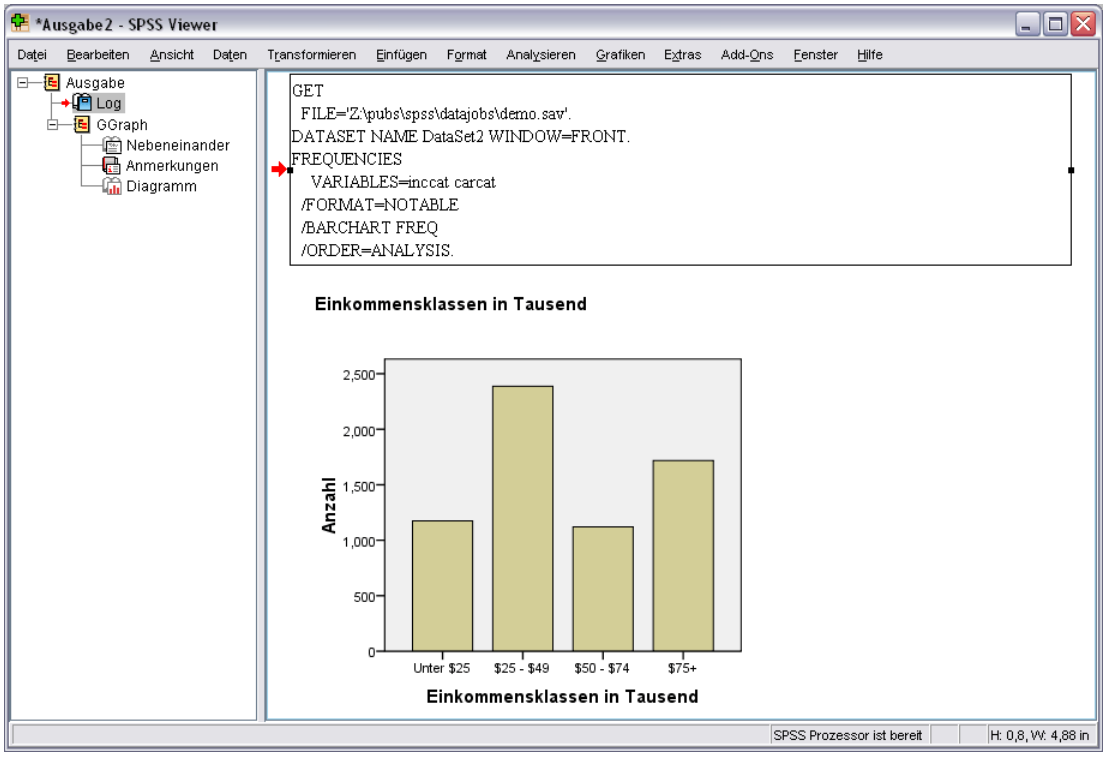
Kopieren von Syntax aus dem Ausgabe-Log

Sie können eine Syntaxdatei erstellen, indem Sie die Befehlssyntax aus dem SPSS-Log kopieren, das im Viewer angezeigt wird. Dazu müssen Sie in den Viewer-Einstellungen (Menü "Bearbeiten", "Optionen", Registerkarte "Viewer") die Option Befehle im Log anzeigen auswählen,

bevor Sie die Analyse ausführen. Jeder Befehl wird dann zusammen mit der Ausgabe der Analyse im Viewer angezeigt.

In einem Syntax-Fenster können Sie die übernommene Syntax ausführen, bearbeiten und in einer Syntaxdatei speichern.

Abbildung 12-2
Befehlssyntax im Log



So kopieren Sie die Syntax aus dem Ausgabe-Log:

- Bevor Sie die Analyse ausführen, wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Optionen...

- Wählen Sie auf der Registerkarte "Viewer" die Option Befehle im Log anzeigen.

  Wenn Sie die Analyse ausführen, werden die SPSS-Befehle für Ihre Auswahl im Dialogfeld im Log aufgezeichnet.

- Öffnen Sie eine vorher gespeicherte Syntaxdatei oder erstellen Sie eine neue. Wählen Sie zum Erstellen einer neuen Syntaxdatei die folgenden Befehle aus den Menüs aus:

  Datei
  Neu
  Syntax

- Doppelklicken Sie im Viewer auf einen Log-Eintrag, um diesen zu aktivieren.

- Wählen Sie den Text aus, der kopiert werden soll.

- Wählen Sie die folgenden Befehle aus den Menüs des Viewers aus:

  Bearbeiten
  Kopieren

- Wählen Sie die folgenden Befehle aus den Menüs in einem Syntax-Fenster aus:

  Bearbeiten
  Einfügen

So führen Sie Befehlssyntax aus:

- Markieren Sie die Befehle, die Sie im Syntax-Fenster ausführen möchten.
- Klicken Sie auf das Schaltfeld "Ausführen" (das nach rechts zeigende Dreieck) auf der Symbolleiste des Syntax-Editors.

  oder

- Wählen Sie einen der Befehle aus dem Menü "Ausführen".
  - **Alles.** Führt alle Befehle im Syntax-Fenster aus.
  - **Auswahl.** Führt die aktuell ausgewählten Befehle aus. Dies umfaßt alle auch nur teilweise markierten Befehle. Wenn keine Auswahl vorgenommen wurde, wird der Befehl ausgeführt, bei dem sich der Cursor gerade befindet.
  - **Aktuell.** Führt den Befehl aus, bei dem sich der Cursor gerade befindet.
  - **Bis Ende.** Führt alle Befehle von der aktuellen Cursorposition bis zum Ende der Befehlssyntax aus.

  Das Schaltfeld "Ausführen" auf der Symbolleiste des Syntaxeditors führt die ausgewählten Befehle aus. Falls Sie keine Auswahl getroffen haben, wird der Befehl, bei dem sich der Cursor gerade befindet, ausgeführt.

  Abbildung 12-3
  Schaltfläche "Ausführen"

Unicode-Syntaxdateien

Im Unicode-Modus ist das Standardformat zum Speichern von Befehlssyntax-Dateien, die während der Sitzung erstellt oder geändert wurden, ebenfalls Unicode (UTF-8). Befehlssyntaxdateien im Unicode-Format können nicht von Versionen von SPSS vor 16.0 gelesen werden. Weitere Informationen zum Unicode-Modus finden Sie unter Optionen: Allgemein auf S. 491.

So speichern Sie eine Syntaxdatei in einem Format, das mit früheren Versionen kompatibel ist:

- Wählen Sie die folgenden Optionen aus den Menüs des Syntaxfensters aus:

  Datei
  Speichern unter

- Wählen Sie im Dialogfeld “Speichern unter” in der Dropdown-Liste “Kodierung” die Option Lokale Kodierung aus. Die lokale Kodierung richtet sich nach dem aktuellen Gebietsschema.

Mehrere Execute-Befehle

Syntax, die aus Dialogfeldern eingefügt bzw. aus dem Protokoll oder Journal kopiert wird, kann Befehle EXECUTE enthalten. Wenn Sie mehrere Befehle über ein Syntax-Fenster ausführen, sind die EXECUTE-Befehle im Allgemeinen überflüssig und können die Leistung herabsetzen, insbesondere bei größeren Datendateien, da bei jedem EXECUTE-Befehl die gesamte Datendatei gelesen wird. Weitere Informationen finden Sie unter dem Befehl EXECUTE in der *Command Syntax Reference* (verfügbar über das Menü “Hilfe” in jedem SPSS-Fenster).

Intervallfunktionen

Eine wichtige Ausnahme stellen Transformationsbefehle dar, die Intervallfunktionen beinhalten. Bei einer Reihe von Transformationsbefehlen ohne trennenden Befehl EXECUTE oder andere Befehle, die die Daten auslesen, werden die Intervallfunktionen unabhängig von der Befehlsreihenfolge nach allen anderen Transformationen berechnet. Beispiel:

```
COMPUTE lagvar=LAG(var1).
COMPUTE var1=var1*2.
```

und

```
COMPUTE lagvar=LAG(var1).
EXECUTE.
COMPUTE var1=var1*2.
```

führen zu höchst unterschiedlichen Ergebnissen für den Wert von *lagvar*, da bei ersterem der transformiert Wert von *var1* verwendet wird und bei letzterem der ursprüngliche Wert.

Kapitel
13

Häufigkeiten

Die Prozedur “Häufigkeiten” stellt Statistiken und grafische Darstellungen für die Beschreibung vieler Variablentypen zur Verfügung. Die Prozedur “Häufigkeiten” ist ein guter Ausgangspunkt für die Betrachtung Ihrer Daten.

Bei Häufigkeitsberichten und Balkendiagrammen können Sie die unterschiedlichen Werte in aufsteigender oder absteigender Reihenfolge anordnen oder die Kategorien nach deren Häufigkeiten ordnen. Der Häufigkeitsbericht kann unterdrückt werden, wenn für eine Variable viele unterschiedliche Werte vorhanden sind. Sie können Diagramme mit Häufigkeiten (die Standardeinstellung) oder Prozentsätzen beschriften.

Beispiel. Wie sind die Kunden eines Unternehmens nach Industriezweigen verteilt? Sie können aus Ihren Ausgabedaten ersehen, dass 37,5 % Ihrer Kunden zu staatlichen Behörden gehören, 24,9 % zu Unternehmen der freien Wirtschaft, 28,1 % zu akademischen Institutionen und 9,4 % zum Gesundheitswesen. Bei stetigen quantitativen Daten wie Verkaufserlösen könnten Sie beispielsweise ersehen, dass sich der durchschnittliche Produktverkauf auf $3.576 bei einer Standardabweichung von $1.078 beläuft.

Statistiken und Diagramme. Häufigkeiten, Prozentsätze, kumulierte Prozentsätze, Mittelwert, Median, Modalwert, Summe, Standardabweichung, Varianz, Spannweite, Minimum und Maximum, Standardfehler des Mittelwerts, Schiefe und Kurtosis (beide mit Standardfehler), Quartile, benutzerdefinierte Perzentile, Balkendiagramme, Kreisdiagramme und Histogramme.

Daten. Verwenden Sie zum Kodieren kategorialer Variablen (nominales oder ordinales Messniveau) numerische Codes oder Strings.

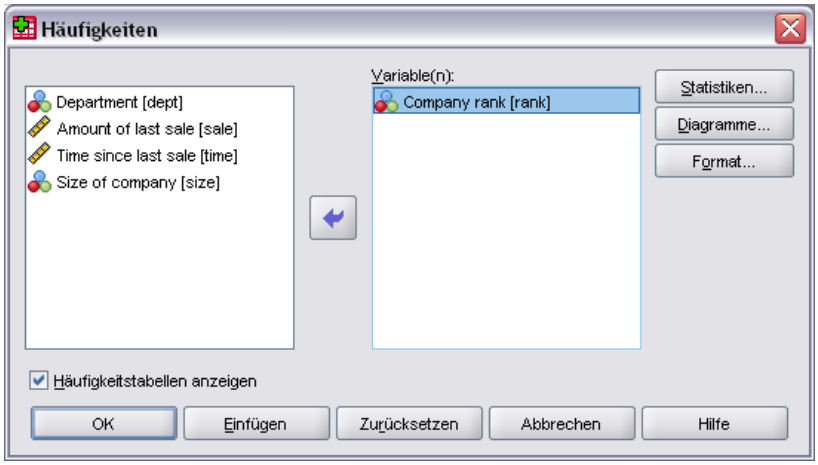
Annahmen. Die Tabellen und Prozentsätze stellen nützliche Beschreibungen für Daten aus allen Verteilungen zur Verfügung, insbesondere für Variablen mit geordneten oder ungeordneten Kategorien. Die meisten der optionalen Auswertungsstatistiken, wie zum Beispiel der Mittelwert und die Standardabweichung, gehen von der Normalverteilung aus und können auf quantitative Variablen mit symmetrischen Verteilungen angewendet werden. Robuste Statistiken, wie zum Beispiel Median, Quartile und Perzentile, sind für quantitative Variablen geeignet, die nur möglicherweise die Annahme erfüllen, dass eine Normalverteilung gilt.

So erstellen Sie Häufigkeitstabellen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Häufigkeiten...

Abbildung 13-1
Hauptdialogfeld von "Häufigkeiten"



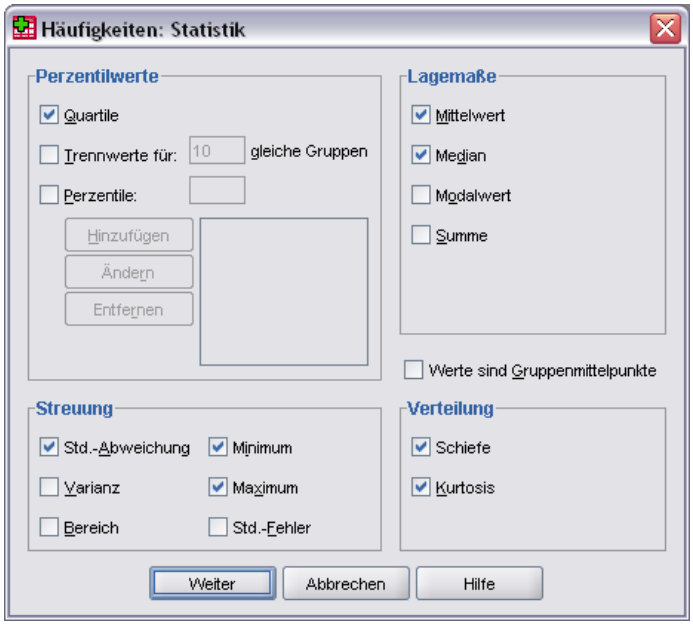
- Wählen Sie mindestens eine kategoriale oder quantitative Variable aus.

Die folgenden Optionen sind verfügbar:

- Deskriptive Statistiken für quantitative Variablen erhalten Sie, indem Sie auf Statistik klicken.
- Balkendiagramme, Kreisdiagramme oder Histogramme erhalten Sie, indem Sie auf Diagramme klicken.
- Sie können die Reihenfolge der angezeigten Ergebnisse ändern, indem Sie auf Format klicken.

Häufigkeiten: Statistik

Abbildung 13-2
Dialogfeld "Häufigkeiten: Statistik"



Perzentilwerte. Dies sind Werte einer quantitativen Variablen, welche die geordneten Daten in Gruppen unterteilen, sodass ein bestimmter Prozentsatz darüber und ein bestimmter Prozentsatz darunter liegt. Quartile (die 25., 50. und 75. Perzentile) unterteilen die Beobachtung in vier gleich große Gruppen. Falls Sie eine gleiche Anzahl von Gruppen wünschen, die von vier abweicht, klicken Sie auf Trennen und geben Sie eine Anzahl für "gleiche Gruppen" ein. Sie können auch individuelle Perzentile festlegen (zum Beispiel das 95. Perzentil, also der Wert, unter dem 95 % der Beobachtungen liegen).

Lagemaße. Statistiken, welche die Lage der Verteilung beschreiben, sind Mittelwert, Median, Modalwert und Summe aller Werte.

- **Mittelwert.** Ein Lagemaß. Das arithmetische Mittel, d. h. die Summe geteilt durch die Anzahl der Fälle.
- **Median.** Wert, über und unter dem jeweils die Hälfte der Fälle liegt; 50. Perzentil. Bei einer geraden Anzahl von Fällen ist der Median der Mittelwert der beiden mittleren Fälle, wenn diese auf- oder absteigend sortiert sind. Der Median ist ein Lagemaß, das gegenüber Ausreißern unempfindlich ist (im Gegensatz zum Mittelwert, der durch wenige extrem niedrige oder hohe Werte beeinflusst werden kann).
- **Modalwert.** Der am häufigsten auftretende Wert. Wenn mehrere Werte gleichermaßen die größte Häufigkeit aufweisen, ist jeder von ihnen ein Modalwert. Die Prozedur "Häufigkeiten" meldet bei mehreren Modalwerten nur den kleinsten.
- **Summe.** Die Summe der Werte über alle Fälle mit nichtfehlenden Werten.

Streuung. Statistiken, welche die Menge an Variation oder die Streubreite in den Daten messen, sind Standardabweichung, Varianz, Spannweite, Minimum, Maximum und Standardfehler des Mittelwerts.

- **Std.abweichung.** Ein Maß für die Streuung um den Mittelwert. Bei einer Normalverteilung liegen 68 % der Fälle im Bereich von einer Standardabweichung um den Mittelwert und 95 % der Fälle im Bereich von zwei Standardabweichungen. Wenn beispielsweise für das Alter der Mittelwert 45 und die Standardabweichung 10 beträgt, liegen bei einer Normalverteilung 95 % der Fälle im Bereich zwischen 25 und 65.
- **Varianz.** Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Maßeinheit der Varianz ist das Quadrat der Maßeinheiten der Variablen.
- **Spannweite.** Die Differenz zwischen den größten und kleinsten Werten einer numerischen Variablen; Maximalwert minus Minimalwert.
- **Minimum.** Der kleinste Wert einer numerischen Variablen.
- **Maximum.** Der größte Wert einer numerischen Variablen.
- **Standardfehler des Mittelwerts.** Ein Maß für die mögliche Variation des Mittelwerts zwischen aus derselben Verteilung stammenden Stichproben. Dieser Wert kann für einen ungefähren Vergleich des beobachteten Mittelwerts mit einem hypothetischen Wert verwendet werden. (Es kann geschlossen werden, dass die beiden Werte unterschiedlich sind, wenn das Verhältnis der Differenz zum Standardfehler kleiner als -2 oder größer als +2 ist.)

Verteilung. Schiefe und Kurtosis sind Statistiken, die Form und Symmetrie der Verteilung beschreiben. Diese Statistiken werden mit ihren Standardfehlern angezeigt.

- **Schiefe.** Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert 0. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (lange rechte Flanke). Eine Verteilung mit einer deutlichen negativen Schiefe läuft nach links lang aus (lange linke Flanke). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler, für eine Abweichung von der Symmetrie spricht.
- **Kurtosis.** Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flanken. Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flanken.

Werte sind Gruppenmittelpunkte. Falls die Werte in den Daten Gruppenmittelpunkte sind (wenn zum Beispiel das Alter aller Personen in den Dreißigern mit dem Wert 35 kodiert ist), wählen Sie diese Option, um den Median und das Perzentil für die ursprünglichen, nicht gruppierten Daten berechnen zu lassen.

Häufigkeiten: Diagramme

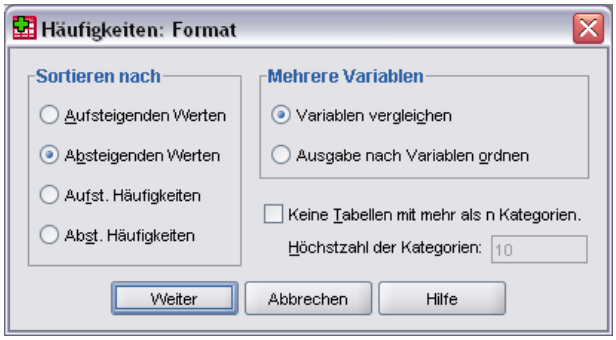
Abbildung 13-3
Dialogfeld "Häufigkeiten: Diagramme"

Diagrammtyp. In einem Kreisdiagramm wird der Anteil der Teile an einem Ganzen angezeigt. Jedes Segment eines Kreisdiagramms entspricht einer durch eine einzelne Gruppenvariable definierten Gruppe. In einem Balkendiagramm wird die Anzahl für jeden unterschiedlichen Wert oder jede unterschiedliche Kategorie als separater Balken angezeigt, wodurch Sie Kategorien visuell vergleichen können. Auch Histogramme enthalten Balken, diese sind jedoch an einer Skala mit gleichen Abständen ausgerichtet. Die Höhe jedes Balkens gibt die Anzahl der Werte einer quantitativen Variablen wieder, die innerhalb des Intervalls liegen. In einem Histogramm werden Form, Mittelpunkt und die Streubreite der Verteilung angezeigt. Eine über das Histogramm gelegte Normalverteilungskurve erleichtert die Beurteilung, ob die Daten normalverteilt sind.

Diagrammwerte. Bei Balkendiagrammen kann die Skalenachse mit Häufigkeiten oder Prozentwerten beschriftet werden.

Häufigkeiten: Format

Abbildung 13-4
Dialogfeld "Häufigkeiten: Format"



Sortieren nach. Die Häufigkeitstabelle kann entsprechend den tatsächlichen Werten der Daten oder entsprechend der Anzahl (Häufigkeit des Vorkommens) dieser Werte geordnet werden. Die Tabelle kann entweder in aufsteigender oder in absteigender Reihenfolge angeordnet werden. Wenn Sie allerdings ein Histogramm oder Perzentile anfordern, wird in der Prozedur "Häufigkeiten" davon ausgegangen, dass die Variable quantitativ ist. Die Werte werden dann in aufsteigender Reihenfolge angezeigt.

Mehrere Variablen. Wenn Sie Statistiktabellen für multiple Variablen erzeugen, können Sie entweder alle Variablen in einer einzigen Tabelle (Variablen vergleichen) oder eine eigene Statistiktabelle für jede Variable (Ausgabe nach Variablen ordnen) anzeigen.

Keine Tabellen mit mehr als n Kategorien. Diese Option verhindert die Anzeige von Tabellen mit mehr als der angegebenen Anzahl von Werten.

Kapitel
14

Deskriptive Statistiken

Mit der Prozedur "Deskriptive Statistiken" werden in einer einzelnen Tabelle univariate Auswertungsstatistiken für verschiedene Variablen angezeigt und standardisierte Werte (*Z*-Werte) errechnet. Variablen können folgendermaßen geordnet werden: nach der Größe ihres Mittelwerts (in aufsteigender oder absteigender Reihenfolge), alphabetisch oder in der Reihenfolge, in der sie ausgewählt wurden (dies ist die Standardeinstellung).

Wenn *Z*-Werte gespeichert werden, werden sie zu den Daten im Daten-Editor hinzugefügt und stehen dann für SPSS-Diagramme, Auflistungen von Daten und Analysen zur Verfügung. Wenn Variablen in verschiedenen Einheiten aufgezeichnet werden (zum Beispiel Bruttoinlandsprodukt pro Kopf der Bevölkerung und Prozentsatz der Alphabetisierung), werden die Variablen durch eine *Z*-Wert-Transformation zur Erleichterung des visuellen Vergleichs auf einer gemeinsamen Skala angeordnet.

Beispiel. Sie zeichnen über mehrere Monate den täglichen Umsatz jedes einzelnen Angestellten der Verkaufsabteilung auf (z. B. ein Eintrag für Herbert, ein Eintrag für Sabine und ein Eintrag für Joachim), sodass jeder Fall in Ihren Daten den täglichen Umsatz jedes Angestellten enthält. Mit der Prozedur "Deskriptive Statistik" wird für Sie jetzt der durchschnittliche Tagesumsatz der einzelnen Angestellten berechnet und das Ergebnis vom höchsten durchschnittlichen Umsatz zum niedrigsten durchschnittlichen Umsatz geordnet.

Statistiken. Stichprobengröße, Mittelwert, Minimum, Maximum, Standardabweichung, Varianz, Spannweite, Summe, Standardfehler des Mittelwerts und Kurtosis und Schiefe mit den Standardfehlern.

Daten. Verwenden Sie numerische Variablen, nachdem Sie diese im Diagramm auf Aufzeichnungsfehler, Ausreißer und Unregelmäßigkeiten in der Verteilung untersucht haben. Die Prozedur "Deskriptive Statistiken" ist für große Dateien (mit Tausenden von Fällen) besonders effektiv.

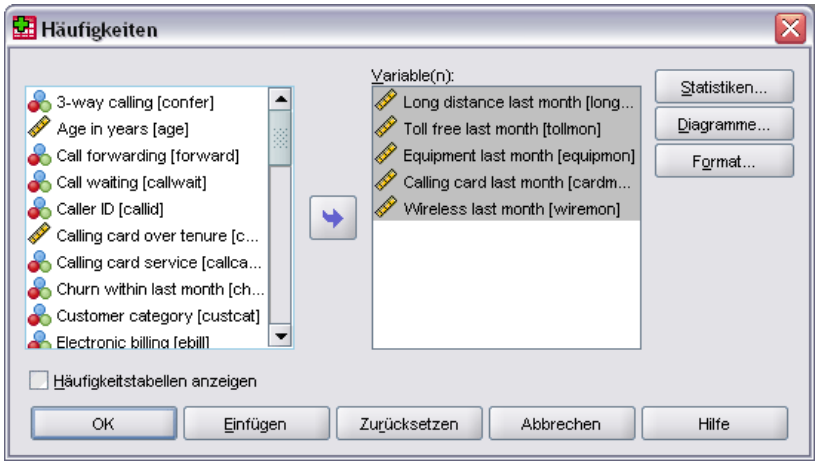
Annahmen. Die meisten verfügbaren Statistiken (einschließlich *Z*-Werte) basieren auf der Annahme, dass die Daten normalverteilt sind, und sind für quantitative Variablen (mit Intervall- oder Verhältnis-Messniveau) mit symmetrischen Verteilungen geeignet. Vermeiden Sie Variablen mit ungeordneten Kategorien oder schiefen Verteilungen. Die Verteilung der *Z*-Werte hat dieselbe Form wie die ursprünglichen Daten; daher bietet das Berechnen von *Z*-Werten keine Abhilfe bei problematischen Daten.

So lassen Sie deskriptive Statistiken berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Deskriptive Statistiken...

Abbildung 14-1
Dialogfeld "Deskriptive Statistik"



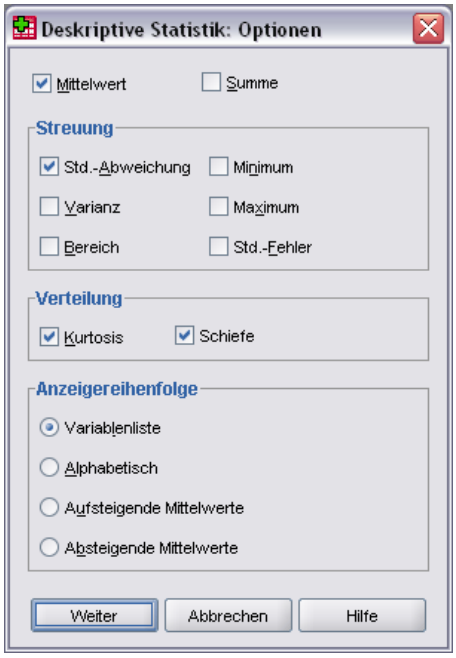
- Wählen Sie mindestens eine Variable aus.

Die folgenden Optionen sind verfügbar:

- Wählen Sie Standardisierte Werte als Variable speichern, um *Z*-Werte als neue Variablen zu speichern.
- Optionale Statistiken und die Reihenfolge der Anzeige steuern Sie, indem Sie auf Optionen klicken.

Deskriptive Statistik: Optionen

Abbildung 14-2
Dialogfeld "Deskriptive Statistik: Optionen"



Mittelwert und Summe. In der Standardeinstellung wird der Mittelwert bzw. das arithmetische Mittel angezeigt.

Streuung. Zu den Statistiken, welche die Streubreite oder die Variation in den Daten messen, gehören Standardabweichung, Varianz, Spannweite, Minimum, Maximum und Standardfehler des Mittelwerts.

- **Std.abweichung.** Ein Maß für die Streuung um den Mittelwert. Bei einer Normalverteilung liegen 68 % der Fälle im Bereich von einer Standardabweichung um den Mittelwert und 95 % der Fälle im Bereich von zwei Standardabweichungen. Wenn beispielsweise für das Alter der Mittelwert 45 und die Standardabweichung 10 beträgt, liegen bei einer Normalverteilung 95 % der Fälle im Bereich zwischen 25 und 65.
- **Varianz.** Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Maßeinheit der Varianz ist das Quadrat der Maßeinheiten der Variablen.
- **Spannweite.** Die Differenz zwischen den größten und kleinsten Werten einer numerischen Variablen; Maximalwert minus Minimalwert.
- **Minimum.** Der kleinste Wert einer numerischen Variablen.
- **Maximum.** Der größte Wert einer numerischen Variablen.
- **Standardfehler des Mittelwerts.** Ein Maß für die mögliche Variation des Mittelwerts zwischen aus derselben Verteilung stammenden Stichproben. Dieser Wert kann für einen ungefähren Vergleich des beobachteten Mittelwerts mit einem hypothetischen Wert verwendet werden. (Es kann geschlossen werden, dass die beiden Werte unterschiedlich sind, wenn das Verhältnis der Differenz zum Standardfehler kleiner als -2 oder größer als +2 ist.)

Verteilung. Kurtosis und Schiefe sind Statistiken, die Form und Symmetrie der Verteilung charakterisieren. Diese Statistiken werden mit ihren Standardfehlern angezeigt.

- **Kurtosis.** Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flanken. Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flanken.
- **Schiefe.** Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert 0. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (lange rechte Flanke). Eine Verteilung mit einer deutlichen negativen Schiefe läuft nach links lang aus (lange linke Flanke). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler, für eine Abweichung von der Symmetrie spricht.

Anzeigereihenfolge. In der Standardeinstellung werden die Variablen in der Reihenfolge angezeigt, in der sie ausgewählt wurden. Sie können Variablen bei Bedarf in alphabetischer Reihenfolge mit aufsteigend oder absteigend geordneten Mittelwerten anzeigen lassen.

Zusätzliche Funktionen beim Befehl DESCRIPTIVES

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Sie können die standardisierten Werte (*Z*-Werte) selektiv für einige Variablen speichern (mit dem Unterbefehl `VARIABLES`).
- Sie können Namen für die neuen Variablen angeben, die die standardisierte Werte enthalten (mit dem Unterbefehl `VARIABLES`).
- Sie können Fälle mit fehlenden Werten in einer beliebigen Variablen aus der Analyse ausschließen (mit dem Unterbefehl `MISSING`).
- Sie können die Variablen in der Anzeige nach dem Wert einer beliebigen Statistik, nicht nur nach dem Mittelwert sortieren (mit dem Unterbefehl `SORT`).

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
15

Explorative Datenanalyse

Mit der Prozedur “Explorative Datenanalyse” werden Auswertungsstatistiken und grafische Darstellungen für alle Fälle oder für separate Fallgruppen erzeugt. Es kann viele Gründe für die Verwendung der Prozedur “Explorative Datenanalyse” geben: Sichten von Daten, Erkennen von Ausreißern, Beschreibung, Überprüfung der Annahmen und Charakterisieren der Unterschiede zwischen Teilgrundgesamtheiten (Fallgruppen). Beim Sichten der Daten können Sie ungewöhnliche Werte, Extremwerte, Lücken in den Daten oder andere Auffälligkeiten erkennen. Durch die explorative Datenanalyse können Sie sich vergewissern, ob die für die Datenanalyse vorgesehenen statistischen Methoden geeignet sind. Die Untersuchung kann ergeben, dass Sie die Daten transformieren müssen, falls die Methode eine Normalverteilung erfordert. Sie können sich stattdessen auch für die Verwendung nichtparametrischer Tests entscheiden.

Beispiel. Betrachten Sie die Verteilung der Lernzeiten für Ratten im Labyrinth mit vier verschiedenen Schwierigkeitsgraden. Zu jeder der vier Gruppen können Sie ablesen, ob die Zeiten annähernd normalverteilt und die vier Varianzen gleich sind. Sie können auch die Fälle mit den fünf längsten und den fünf kürzesten Zeiten bestimmen. Sie können die Verteilung der Lernzeiten für jede Gruppe mit Boxplots und Stengel-Blatt-Diagrammen grafisch auswerten.

Statistiken und Diagramme. Mittelwert, Median, 5% getrimmtes Mittel, Standardfehler, Varianz, Standardabweichung, Minimum, Maximum, Spannweite, interquartiler Bereich, Schiefe und Kurtosis und deren Standardfehler, Konfidenzintervall für den Mittelwert (und angegebenes Konfidenzniveau), Perzentile, M-Schätzer nach Huber, Andrew-Wellen-Schätzer, M-Schätzer nach Hampel, Tukey-Biweight-Schätzer, die fünf größten und die fünf kleinsten Werte, die Kolmogorov-Smirnov-Statistik mit Lilliefors-Signifikanzniveau zum Prüfen der Normalverteilung und die Shapiro-Wilk-Statistik. Boxplots, Stengel-Blatt-Diagramme, Histogramme, Normalverteilungsdiagramme und Diagramme der Streubreite gegen das mittlere Niveau mit Levene-Test und Transformationen.

Daten. Die Prozedur “Explorative Datenanalyse” kann für quantitative Variablen (mit Intervall- oder Verhältnis-Messniveau) verwendet werden. Eine Faktorvariable (zum Aufteilen der Daten in Fallgruppen) muss eine sinnvolle Anzahl von unterschiedlichen Werten (Kategorien) enthalten. Diese Werte können kurze Strings oder numerische Werte sein. Die Fallbeschriftungsvariable, die für die Beschriftung von Ausreißern in Boxplots verwendet wird, kann ein kurzer String, ein langer String (die ersten 15 Byte) oder numerisch sein.

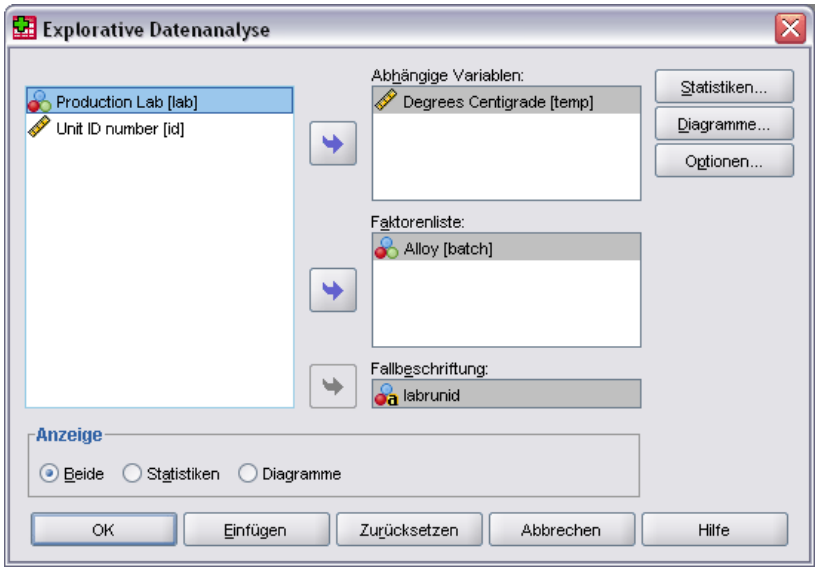
Annahmen. Ihre Daten müssen nicht symmetrisch oder normalverteilt sein.

So führen Sie eine explorative Datenanalyse aus:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Explorative Datenanalyse

Abbildung 15-1
Dialogfeld "Explorative Datenanalyse"



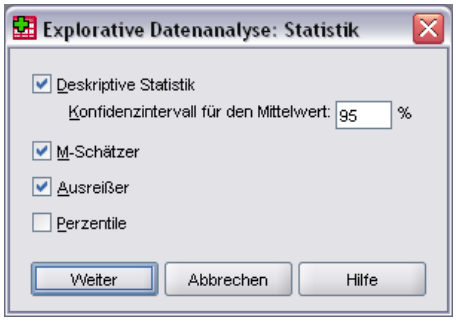
▶ Wählen Sie eine oder mehrere abhängige Variablen aus.

Die folgenden Optionen sind verfügbar:

- Auswählen einer oder mehrerer Faktorvariablen, mit deren Werten Fallgruppen definiert werden.
- Auswählen einer Identifizierungsvariablen für die Beschriftung von Fällen.
- Zugriff auf robuste Schätzer, Ausreißer, Perzentile und Häufigkeitstabellen erhalten Sie, indem Sie auf Statistik klicken.
- Zugriff auf Histogramme, Normalverteilungsdiagramme und Tests sowie Diagramme der Streubreite gegen das mittlere Niveau mit Levene-Statistik erhalten Sie, indem Sie auf Diagramme klicken.
- Sie können die Behandlung fehlender Werte festlegen, indem Sie auf Optionen klicken.

Explorative Datenanalyse: Statistik

Abbildung 15-2
Dialogfeld "Explorative Datenanalyse: Statistik"



Deskriptive Statistiken. In der Standardeinstellung werden Lage- und Streuungsmaße angezeigt. Mit den Lagemaßen wird die Lage der Verteilung angegeben. Dazu gehören Mittelwert, Median und 5% getrimmtes Mittel. Mit den Maßen für Streuung werden Unähnlichkeiten der Werte angezeigt. Diese umfassen Standardfehler, Varianz, Standardabweichung, Minimum, Maximum, Spannweite und den Interquartilbereich. Die beschreibenden Statistiken enthalten auch Maße der Verteilungsform. Schiefe und Kurtosis werden mit den jeweiligen Standardfehlern angezeigt. Das 95%-Konfidenzintervall für den Mittelwert wird ebenfalls angezeigt. Sie können auch ein anderes Konfidenzniveau angeben.

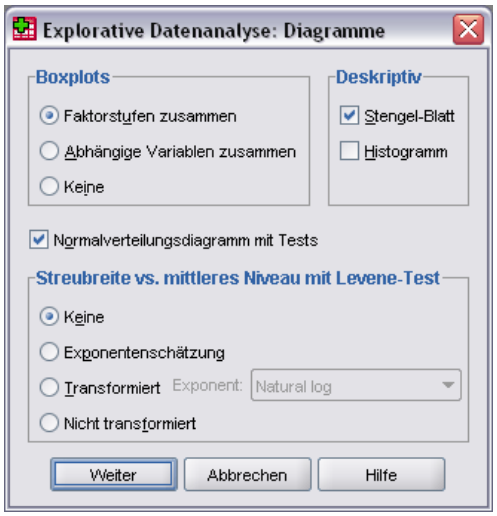
M-Schätzer. Robuste Alternativen zu Mittelwert und Median der Stichprobe zum Schätzen der Lage. Die berechneten Schätzer unterscheiden sich in den Gewichtungen, die sie den Fällen zuweisen. M-Schätzer nach Huber, Andrew-Wellen-Schätzer, M-Schätzer nach Hampel und Tukey-Biweight-Schätzer werden angezeigt.

Ausreißer. Hier werden die fünf größten und die fünf kleinsten Werte mit Fallbeschriftungen angezeigt.

Perzentile. Hier werden die Werte für die 5., 10., 25., 50., 75., 90. und 95. Perzentile angezeigt.

Explorative Datenanalyse: Diagramme

Abbildung 15-3
Dialogfeld "Explorative Datenanalyse: Diagramme"



Boxplots. Mit diesen Optionen legen Sie fest, wie Boxplots bei mehr als einer abhängigen Variablen angezeigt werden. Mit Faktorstufen zusammen wird eine getrennte Anzeige für jede abhängige Variable erzeugt. In einer Anzeige werden Boxplots für alle durch eine Faktorvariable definierten Gruppen angezeigt. Mit Abhängige Variablen zusammen wird für jede durch eine Faktorvariable definierte Gruppe eine getrennte Anzeige erzeugt. In einer Anzeige werden Boxplots für alle abhängigen Variablen in einer Anzeige nebeneinander dargestellt. Diese Anzeige ist insbesondere nützlich, wenn verschiedene Variablen ein einziges, zu unterschiedlichen Zeiten gemessenes Merkmal darstellen.

Deskriptive Statistik. Im Gruppenfeld "Deskriptive Statistik" können Sie Stengel-Blatt-Diagramme und Histogramme auswählen.

Normalverteilungsdiagramme mit Tests. Hier werden Normalverteilungsdiagramme und trendbereinigte Normalverteilungsdiagramme angezeigt. Die Kolmogorov-Smirnov-Statistik mit einem Signifikanzniveau nach Lilliefors für den Test auf Normalverteilung wird angezeigt. Bei Angabe von nichtganzzahligen Gewichtungen wird die Shapiro-Wilk-Statistik berechnet, wenn die gewichtete Stichprobengröße zwischen 3 und 50 liegt. Bei keinen oder ganzzahligen Gewichtungen wird die Statistik berechnet, wenn die gewichtete Stichprobengröße zwischen 3 und 5,000 liegt.

Streubreite vs. mittleres Niveau mit Levene-Test. Hiermit legen Sie fest, wie Daten für Diagramme der Streubreite versus mittleres Niveau transformiert werden. Für alle Diagramme der Streubreite versus mittleres Niveau werden die Steigung der Regressionsgeraden und der Levene-Test auf Homogenität der Varianz angezeigt. Wenn Sie eine Transformation auswählen, liegen dem Levene-Test die transformierten Daten zugrunde. Wenn keine Faktorvariable ausgewählt wurde, werden keine Diagramme der Streubreite versus mittleres Niveau erstellt. Mit der Exponentenschätzung wird ein Diagramm der natürlichen Logarithmen des Interquartilbereichs über die natürlichen Logarithmen des Medians für alle Zellen sowie eine Schätzung der Potenztransformation zum Erreichen gleicher Varianzen in den Zellen angefordert. Mit

Diagrammen der Streubreite versus mittleres Niveau lässt sich der Exponent für Transformationen bestimmen, mit denen über Gruppen hinweg eine höhere Stabilität (höhere Gleichförmigkeit) der Varianzen erreicht wird. Mit Transformiert können Sie einen alternativen Exponenten auswählen, eventuell gemäß der Empfehlung der Exponentenschätzung, und Diagramme der transformierten Daten erzeugen. Der Interquartilbereich und der Median der transformierten Daten werden grafisch dargestellt. Mit Nicht transformiert werden Diagramme der Rohdaten erstellt. Dies entspricht einer Transformation mit einem Exponenten gleich 1.

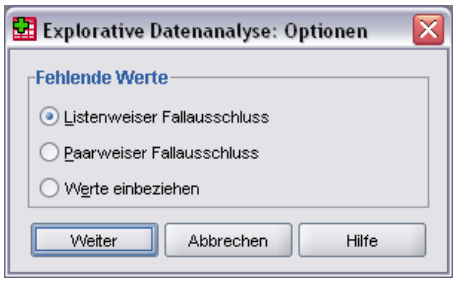
Explorative Datenanalyse: Potenztransformationen

Dies sind die Potenztransformationen für Diagramme der Streubreite versus mittleres Niveau. Für die Transformation von Daten muss ein Exponent ausgewählt werden. Sie können eine der folgenden Möglichkeiten wählen:

- **Natürlicher Logarithmus.** Transformation mit natürlichem Logarithmus. Dies ist die Standardeinstellung.
- **1/Quadratwurzel.** Zu jedem Datenwert wird der reziproke Wert der Quadratwurzel berechnet.
- **Reziprok.** Der reziproke Wert jedes Datenwerts wird berechnet.
- **Quadratwurzel.** Die Quadratwurzel jedes Datenwerts wird berechnet.
- **Quadratisch.** Jeder Datenwert wird quadriert.
- **Kubisch.** Es wird die dritte Potenz jedes Datenwerts errechnet.

Explorative Datenanalyse: Optionen

Abbildung 15-4
Dialogfeld "Explorative Datenanalyse: Optionen"



Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für abhängige Variablen oder Faktorvariablen werden aus allen Analysen ausgeschlossen. Dies ist die Standardeinstellung.
- **Paarweiser Fallausschluss.** Fälle ohne fehlenden Werte für Variablen in einer Gruppe (Zelle) werden in die Analyse dieser Gruppe einbezogen. Der Fall kann fehlende Werte für Variablen enthalten, die in anderen Gruppen verwendet werden.
- **Werte einbeziehen.** Fehlende Werte für Faktorvariablen werden als gesonderte Kategorie behandelt. Die gesamte Ausgabe wird auch für diese zusätzliche Kategorie erstellt. Häufigkeitstabellen enthalten Kategorien für fehlende Werte. Fehlende Werte für Faktorvariablen werden aufgenommen, jedoch als fehlend beschriftet.

Zusätzliche Funktionen beim Befehl EXAMINE

In der Prozedur "Explorative Datenanalyse" wird die Befehlssyntax von EXAMINE verwendet. Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Anfordern von Ausgaben und Diagrammen für Gesamtsummen neben den Ausgaben und Diagrammen für Gruppen, die durch die Faktorvariablen definiert wurden (mit dem Unterbefehl TOTAL).
- Angeben einer gemeinsamen Skala für eine Gruppe von Boxplots (mit dem Unterbefehl SCALE).
- Angeben von Interaktionen der Faktorvariablen (mit dem Unterbefehl VARIABLES).
- Angeben von anderen Perzentilen als in der Standardeinstellung (mit dem Unterbefehl PERCENTILES).
- Berechnen der Perzentile nach fünf Methoden (mit dem Unterbefehl PERCENTILES).
- Angeben einer Potenztransformation für Diagramme der Streubreite gegen das mittlere Niveau (mit dem Unterbefehl PLOT).
- Angeben der Anzahl von Extremwerten, die angezeigt werden sollen (mit dem Unterbefehl STATISTICS).
- Angeben der Parameter für die M-Schätzer, den robusten Schätzern der Lage (mit dem Unterbefehl MESTIMATORS).

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
16

Kreuztabellen

Mit der Prozedur “Kreuztabellen” erzeugen Sie Zweifach- und Mehrfach-Tabellen. Es stehen eine Vielzahl von Tests und Zusammenhangsmaßen für Zweifach-Tabellen zur Verfügung. Welcher Test oder welches Maß verwendet wird, hängt von der Struktur der Tabelle ab und davon, ob die Kategorien geordnet sind.

Statistiken und Zusammenhangsmaße für Kreuztabellen werden nur für Zweifach-Tabellen berechnet. Wenn Sie eine Zeile, eine Spalte und einen Schichtfaktor (Kontroll-Variable) festlegen, wird von der Prozedur “Kreuztabelle” eine separate Ausgabe mit der entsprechenden Statistik sowie den Maßen für jeden Wert des Schichtfaktors (oder eine Kombination der Werte für zwei oder mehrere Kontroll-Variablen) angezeigt. Wenn zum Beispiel *Geschlecht* ein Schichtfaktor für eine Tabelle ist, wobei *verheiratet* (Ja, Nein) gegenüber *Leben* (ist das Leben aufregend, Routine oder langweilig) untersucht wird, werden die Ergebnisse für eine Zweifach-Tabelle für weibliche Personen getrennt von den männlichen berechnet und als aufeinander folgende separate Ausgaben gedruckt.

Beispiel. Wie groß ist die Wahrscheinlichkeit, dass mit den Kunden aus kleineren Unternehmen beim Verkauf von Dienstleistungen (zum Beispiel Weiterbildung und Beratung) ein größerer Gewinn erzielt wird als mit den Kunden aus größeren Unternehmen? Einer Kreuztabelle könnten Sie möglicherweise entnehmen, dass die Mehrheit der kleinen Unternehmen (mit mehr als 500 Angestellten) beim Verkauf von Dienstleistungen einen hohen Gewinn erzielt, während die meisten großen Unternehmen (mit mehr als 2,500 Angestellten) dabei nur niedrige Gewinne erzielen.

Statistiken und Zusammenhangsmaße. Pearson-Chi-Quadrat, Likelihood-Quotienten-Chi-Quadrat, Zusammenhangstest linear-mit-linear, Exakter Test nach Fisher, korrigiertes Chi-Quadrat nach Yates, Pearson-*r*, Spearman-Rho, Kontingenzkoeffizient, Phi, Cramér-*V*, symmetrische und asymmetrische Lambdas, Goodman-und-Kruskal-Tau, Unsicherheitskoeffizient, Gamma, Somer-*d*, Kendall-Tau-*b*, Kendall-Tau-*c*, Eta-Koeffizient, Cohen-Kappa, relativer Risikoschätzer, Quotenverhältnis, McNemar-Test, Cochran- und Mantel-Haenszel-Statistik.

Daten. Um die Kategorien der Tabellenvariablen zu definieren, verwenden Sie Werte einer numerischen Variablen oder einer String-Variablen (maximal 8 Byte). Zum Beispiel können Sie die Daten für *Geschlecht* als 1 und 2 oder als *männlich* und *weiblich* kodieren.

Annahmen. Einige Statistiken und Maße setzen geordnete Kategorien (Ordinal-Daten) oder quantitative Werte (Intervall- oder Verhältnisdaten) voraus, wie bereits im Abschnitt über Statistiken erläutert wurde. Andere sind zulässig, wenn die Tabellenvariablen über ungeordnete Kategorien verfügen (Nominal-Daten). Für Statistiken, die auf Chi-Quadrat basieren (Phi, Cramér-*V*, Kontingenzkoeffizient), sollten die Daten durch eine Zufallsstichprobe aus einer multinomialen Verteilung bezogen werden.

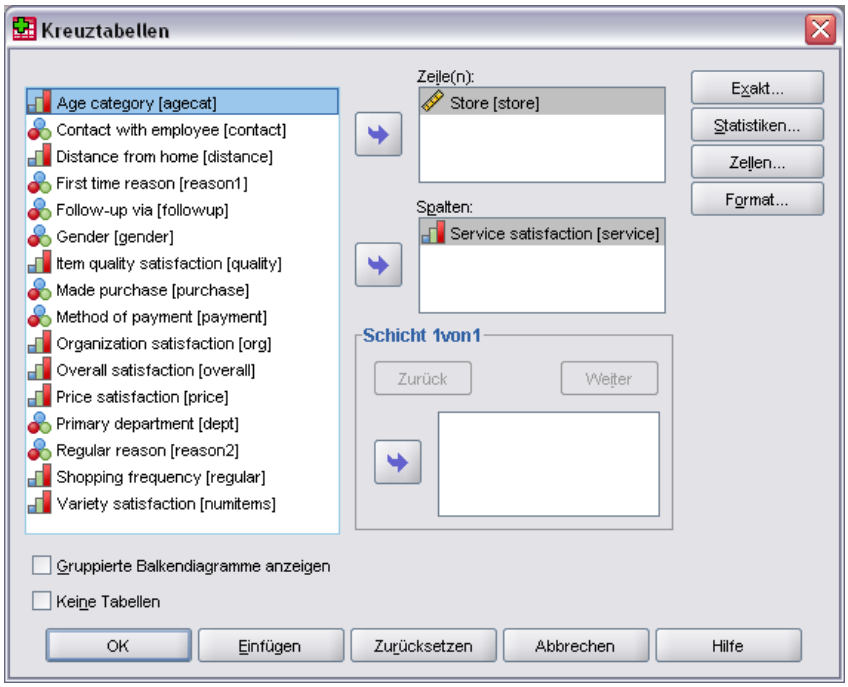
Anmerkung: Bei ordinalen Variablen kann es sich um numerische Codes für Kategorien (z. B. 1 = *schwach*, 2 = *mittel*, 3 = *stark*) oder um String-Werte handeln. Die alphabetische Ordnung der String-Werte gibt dabei die Reihenfolge der Kategorien vor. Bei einer String-Variablen mit den Werten *Schwach*, *Mittel* und *Stark* werden die Kategorien beispielsweise in der Reihenfolge *Mittel*, *Schwach*, *Stark* und somit falsch angeordnet. Im allgemeinen ist die Verwendung von numerischem Code für ordinale Daten günstiger.

So lassen Sie Kreuztabellen berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Kreuztabellen...

Abbildung 16-1
Dialogfeld "Kreuztabellen"



▶ Wählen Sie eine oder mehrere Zeilenvariablen und eine oder mehrere Spaltenvariablen aus.

Die folgenden Optionen sind verfügbar:

- Eine oder mehrere Kontroll-Variablen auswählen.
- Tests und Zusammenhangsmaße der Zweifach-Tabellen oder Untertabellen erhalten Sie, indem Sie auf Statistik klicken.
- Informationen zu beobachteten und erwarteten Werten, Prozentsätzen und Residuen erhalten Sie, indem Sie auf Zellen klicken.
- Durch Klicken auf Format können Sie die Reihenfolge der Kategorien festlegen.

Kreuztabellenschichten

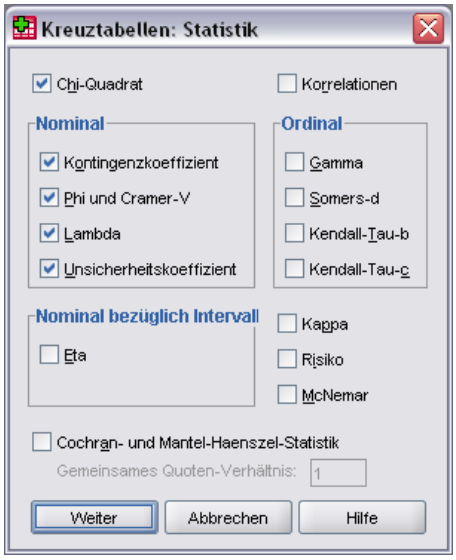
Wenn Sie eine oder mehrere Schichtvariablen auswählen, wird für jede Kategorie jeder Schichtvariablen (Kontroll-Variablen) jeweils eine Kreuztabelle erzeugt. Wenn Sie zum Beispiel über eine Zeilenvariable, eine Spaltenvariable und eine Schichtvariable mit zwei Kategorien verfügen, erhalten Sie eine Zweifach-Tabelle für jede Kategorie der Schichtvariablen. Um eine weitere Schicht von Kontroll-Variablen anzulegen, klicken Sie auf Weiter. Untertabellen werden für jede Kombination von Kategorien für jede Variable der ersten Schicht, jeder Variable der zweiten Schicht und so weiter erzeugt. Wenn Statistiken und Zusammenhangsmaße angefordert werden, treffen diese nur auf Zweifach-Untertabellen zu.

Kreuztabellen: Gruppierte Balkendiagramme

Gruppierte Balkendiagramme anzeigen. Mit einem gruppierten Balkendiagramm können Sie Ihre Daten leichter nach Gruppen von Fällen auswerten. Für jeden Wert der Variablen, der von Ihnen unter Zeilen festgelegt wurde, wird eine Gruppe von Balken erzeugt. Die Balken in jedem Cluster werden durch die unter Spalten angegebene Variable definiert. Für jeden Wert dieser Variablen steht Ihnen ein Set unterschiedlich farbiger oder gemusterter Balken zur Verfügung. Wenn Sie unter Zeilen oder Spalten mehr als eine Variable angeben, wird für jede Kombination von zwei Variablen ein gruppiertes Balkendiagramm erzeugt.

Kreuztabellen: Statistik

Abbildung 16-2
Dialogfeld "Kreuztabellen: Statistik"



Chi-Quadrat. Für Tabellen mit zwei Zeilen und zwei Spalten wählen Sie Chi-Quadrat aus, um das Pearson-Chi-Quadrat, das Likelihood-Quotienten-Chi-Quadrat, den exakten Test nach Fisher und das korrigierte Chi-Quadrat nach Yates (Kontinuitätskorrektur) zu berechnen. Für 2 × 2-Tabellen wird der exakte Test nach Fisher berechnet, wenn eine Tabelle, die nicht aus fehlenden Zeilen

oder Spalten einer größeren Tabelle entstanden ist, eine Zelle mit einer erwarteten Häufigkeit von weniger als 5 enthält. Für alle anderen 2 × 2-Tabellen wird das korrigierte Chi-Quadrat nach Yates berechnet. Für Tabellen mit einer beliebigen Anzahl von Zeilen und Spalten wählen Sie Chi-Quadrat aus, um das Pearson-Chi-Quadrat und das Likelihood-Quotienten-Chi-Quadrat zu berechnen. Wenn beide Tabellenvariablen quantitativ sind, ergibt Chi-Quadrat den Zusammenhangstest linear-mit-linear.

Korrelationen. Für Tabellen, in denen sowohl Zeilen als auch Spalten geordnete Werte enthalten, ergeben die Korrelationen den Korrelationskoeffizienten nach Spearman, also Rho (nur numerische Daten). Der Korrelationskoeffizient nach Spearman ist ein Zusammenhangsmaß zwischen den Rangordnungen. Wenn beide Tabellenvariablen (Faktoren) quantitativ sind, ergibt sich unter Korrelationen der Korrelationskoeffizient nach Pearson, *r*, der ein Maß für den linearen Zusammenhang zwischen den Variablen darstellt.

Nominal. Für nominale Daten (ohne implizierte Reihenfolge, wie beispielsweise katholisch, protestantisch, jüdisch) können Sie Kontingenzkoeffizient, Phi (Koeffizient) und Cramér'-V, Lambda (symmetrische und asymmetrische Lambdas sowie Goodman-und-Kruskal-Tau) und Unsicherheitskoeffizient auswählen.

- **Kontingenzkoeffizient.** Ein auf der Chi-Quadrat-Statistik basierendes Zusammenhangsmaß. Dieser Koeffizient liegt immer zwischen 0 und 1, wobei 0 angibt, dass kein Zusammenhang zwischen Zeilen- und Spaltenvariable besteht und Werte nahe 1 auf einen starken Zusammenhang zwischen den Variablen hindeuten. Der maximale Wert hängt von der Anzahl der Zeilen und Spalten in der Tabelle ab.
- **Phi und Cramer-V.** Phi ist ein auf der Chi-Quadrat-Statistik basierendes Zusammenhangsmaß. Es ergibt sich als Wurzel aus dem Quotienten aus Chi-Quadrat und dem Stichprobenumfang. Cramer-V ist ebenfalls ein Zusammenhangsmaß auf der Basis der Chi-Quadrat-Statistik.
- **Lambda.** Ein Zusammenhangsmaß für die proportionale Fehlerreduktion, wenn Werte der unabhängigen Variablen zur Vorhersage von Werten der abhängigen Variablen verwendet werden. Der Wert 1 bedeutet, dass die abhängige Variable durch die unabhängige Variable vollständig vorhergesagt werden kann. Der Wert 0 bedeutet, dass die Vorhersage der abhängigen Variablen durch die unabhängige Variable nicht unterstützt wird.
- **Unsicherheitskoeffizient.** Ein Zusammenhangsmaß, das die proportionale Fehlerreduktion angibt, wenn Werte einer Variablen zur Vorhersage von Werten der anderen Variablen verwendet werden. Ein Wert von 0,83 gibt z. B. an, dass die Kenntnis einer Variablen den Fehler bei der Vorhersage der Werte der anderen Variablen um 83 % reduziert. Das Programm berechnet beide Versionen des Unsicherheitskoeffizienten, die symmetrische und die asymmetrische.

Ordinal. Für Tabellen, in welchen die Zeilen und Spalten geordnete Werte enthalten, wählen Sie Gamma (nullte Ordnung für Zweifach-Tabellen und bedingt für Dreifach- bis Zehnfach-Tabellen), Kendall-Tau-b und Kendall-Tau-c aus. Zur Vorhersage von Spaltenkategorien auf der Grundlage von Zeilenkategorien wählen Sie Somers-d aus.

- **Gamma.** Ein symmetrisches Zusammenhangsmaß für zwei ordinalskalierte Variablen, dessen Wertebereich zwischen -1 und +1 liegt. Werte nahe bei -1 oder +1 weisen auf einen starken Zusammenhang zwischen den Variablen hin. Werte nahe 0 stehen für einen schwachen oder fehlenden Zusammenhang. Für Tabellen mit zwei Variablen werden Gamma-Werte nullter Ordnung angezeigt. Für Tabellen mit drei oder mehr Variablen werden bedingte Gamma-Werte angezeigt.

- **Somers-d.** Ein Zusammenhangsmaß für zwei ordinale Variablen, dessen Wertebereich zwischen -1 und +1 liegt. Werte, die betragsmäßig nahe bei 1 liegen, geben einen starken Zusammenhang zwischen den beiden Variablen an, Werte nahe 0 einen schwachen oder fehlenden Zusammenhang. Somers-d ist eine asymmetrische Erweiterung von Gamma. Der Unterschied liegt in der Einbeziehung der Anzahl von Paaren, die keine Bindungen in der unabhängigen Variablen aufweisen. Eine symmetrische Version dieser Statistik wird ebenfalls berechnet.
- **Kendall-Tau-b.** Ein nichtparametrisches Korrelationsmaß für ordinale Variablen oder Ränge, das Bindungen berücksichtigt. Das Vorzeichen des Koeffizienten gibt die Richtung des Zusammenhangs an und sein Betrag die Stärke; dabei entsprechen betragsmäßig größere Werte einem stärkeren Zusammenhang. Die möglichen Werte liegen im Bereich von -1 und 1, ein Wert von -1 oder +1 ergibt sich jedoch nur aus quadratischen Tabellen.
- **Kendall-Tau-c.** Ein nichtparametrisches Zusammenhangsmaß für ordinale Variablen, das Bindungen ignoriert. Das Vorzeichen des Koeffizienten gibt die Richtung des Zusammenhangs an und sein Betrag die Stärke; dabei entsprechen betragsmäßig größere Werte einem stärkeren Zusammenhang. Die möglichen Werte liegen im Bereich von -1 und 1, ein Wert von -1 oder +1 ergibt sich jedoch nur aus quadratischen Tabellen.

Nominal bezüglich Intervall. Wenn eine Variable kategorial und eine andere quantitativ ist, wählen Sie Eta aus. Die kategoriale Variable muss numerisch kodiert sein.

- **Eta.** Ein Zusammenhangsmaß, das zwischen 0 und 1 liegt; dabei steht 0 für fehlenden Zusammenhang zwischen den Zeilen- und Spaltenvariablen und Werte nahe bei 1 geben einen starken Zusammenhang an. Eta ist geeignet für eine intervallskalierte abhängige Variable (z. B. Einkommen) und eine unabhängige Variable mit einer begrenzten Anzahl von Kategorien (z. B. Geschlecht). Es werden zwei Eta-Werte berechnet: der eine behandelt die Zeilenvariablen und der andere die Spaltenvariable als intervallskalierte Variable.

Kappa. Der Cohen-Kappa-Koeffizient misst die Übereinstimmung zwischen den Beurteilungen zweier Prüfer, wenn beide dasselbe Objekt bewerten. Der Wert 1 bedeutet perfekte Übereinstimmung. Der Wert 0 bedeutet, dass die Übereinstimmung nicht über das zufallsbedingte Maß hinausgeht. Kappa ist nur für Tabellen verfügbar, in denen beide Variablen die gleiche Anzahl von Kategorien und gleiche Kategorienwerte (Ausprägungen) aufweisen.

Risiko. Ein Maß, das bei 2 x 2-Tabellen die Stärke des Zusammenhangs zwischen dem Vorhandensein eines Faktors und dem Auftreten eines Ereignisses misst. Wenn das Konfidenzintervall für die Statistik den Wert 1 enthält, ist nicht anzunehmen, dass zwischen Faktor und Ereignis ein Zusammenhang besteht. Das Quotenverhältnis (Odds Ratio) kann als Schätzer für das relative Risiko verwendet werden, wenn der Faktor selten auftritt.

McNemar. Ein nichtparametrischer Test für zwei verbundene dichotome Variablen. Prüft unter Verwendung der Chi-Quadrat-Verteilung, ob Änderungen bei den Antworten vorliegen. Dieser Test ist für das Erkennen von Änderungen bei Antworten nützlich, die durch experimentelle Einflussnahme in so genannten "Vorher-und-nachher-Designs" entstanden sind. Bei größeren quadratischen Tabellen wird der McNemar-Bowker-Test auf Symmetrie ausgegeben.

Cochran- und Mantel-Haenszel-Statistik. Die Cochran- und die Mantel-Haenszel-Statistik können verwendet werden, um auf Unabhängigkeit zwischen einer dichotomen Faktorvariablen und einer dichotomen Response-Variablen zu testen, und zwar in Abhängigkeit von einem Kovariatenmuster, das durch eine oder mehrere Schichtvariablen (Kontrollvariablen) definiert

wird. Beachten Sie, dass andere Statistiken schichtenweise berechnet werden, die Cochran- und die Mantel-Haenszel-Statistik dagegen einmal für alle Schichten berechnet werden.

Kreuztabellen: Zellen anzeigen

Abbildung 16-3
Dialogfeld "Kreuztabellen: Zellen anzeigen"

Um Sie beim Erkennen von Mustern in den Daten zu unterstützen, die zu einem signifikanten Chi-Quadrat-Test beitragen, zeigt die Prozedur "Kreuztabellen" die erwarteten Häufigkeiten und drei Typen von Residuen (Abweichungen) an, welche die Differenz zwischen beobachteten und erwarteten Häufigkeiten messen. Jede Zelle der Tabelle kann jede Kombination von ausgewählten Häufigkeiten, Prozentzahlen und Residuen enthalten.

Häufigkeiten. Die Anzahl der Fälle, die tatsächlich beobachtet, und die Anzahl der Fälle, die erwartet werden, wenn die Zeilen- und Spaltenvariablen voneinander unabhängig sind.

Prozentwerte. Die Prozentwerte können horizontal in den Zeilen oder vertikal in den Spalten addiert werden. Der prozentuale Anteil der Gesamtanzahl der Fälle, die in einer Tabelle dargestellt werden (eine Schicht), ist ebenfalls verfügbar.

Residuen. Einfache nicht standardisierte Residuen geben die Differenz zwischen den beobachteten und erwarteten Werten wieder. Standardisierte und korrigierte standardisierte Residuen sind ebenfalls verfügbar.

- **Nicht standardisiert.** Die Differenz zwischen einem beobachteten Wert und dem erwarteten Wert. Der erwartete Wert ist die Anzahl von Fällen, die man in einer Zelle erwarten würde, wenn kein Zusammenhang zwischen den beiden Variablen bestünde. Ein positives Residuum zeigt an, dass in der Zelle mehr Fälle vorliegen, als dies der Fall wäre, wenn die Zeilen- und Spaltenvariable unabhängig wären.

- **Standardisiert.** Der Quotient aus dem Residuum und einem Schätzer seiner Standardabweichung. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- **Korrigiert standardisiert.** Der Quotient aus dem Residuum einer Zelle (beobachteter Wert minus erwarteter Wert) und dessen geschätztem Standardfehler. Das resultierende standardisierte Residuum wird in Einheiten der Standardabweichung über oder unter dem Mittelwert angegeben.

Nichtganzzahlige Gewichtungen. Bei den Zellhäufigkeiten handelt es sich normalerweise um ganzzahlige Werte, da sie für die Anzahl der Fälle in den einzelnen Zellen stehen. Wenn jedoch die Datendatei derzeit mit einer Gewichtungsvariablen mit Bruchzahlenwerten (z. B. 1,25) gewichtet ist, können die Zellhäufigkeiten ebenfalls Bruchwerte sein. Sie können die Werte vor oder nach der Berechnung der Zellhäufigkeiten abschneiden oder runden oder sowohl für die Tabellenanzeige als auch für statistische Berechnungen gebrochene Zellhäufigkeiten verwenden.

- **Anzahl in den Zellen runden.** Fallgewichte werden verwendet, wie gegeben, aber die akkumulierten Gewichte für die Zellen werden gerundet, bevor Statistiken berechnet werden.
- **Anzahl in den Zellen stutzen.** Fallgewichte werden verwendet, wie gegeben, aber die addierten Gewichte für die Zellen werden auf den ganzzahligen Anteil gestutzt, bevor Statistiken berechnet werden.
- **Fallgewichte runden.** Fallgewichte werden gerundet, bevor sie verwendet werden.
- **Fallgewichte stutzen.** Fallgewichte werden auf den ganzzahligen Anteil gestutzt, bevor sie verwendet werden.
- **Keine Korrekturen.** Fallgewichte werden verwendet wie gegeben und auch nicht ganzzahlige Zellhäufigkeiten werden verwendet. Wenn jedoch exakte Statistiken (verfügbar mit dem Modul "Exakte Tests") angefordert werden, dann werden die akkumulierten Gewichte in den Zellen entweder auf den ganzzahligen Anteil gestutzt oder gerundet, bevor die Statistiken für exakte Tests berechnet werden.

Kreuztabellen: Tabellenformat

Abbildung 16-4
Dialogfeld "Kreuztabellen: Tabellenformat"

Sie können Zeilen in aufsteigender oder absteigender Reihenfolge der Werte der Zeilenvariablen anordnen.

Kapitel
17

Zusammenfassen

Mit der Prozedur “Zusammenfassen” werden Untergruppenstatistiken für Variablen innerhalb der Kategorien einer oder mehrerer Gruppenvariablen berechnet. Alle Ebenen der Gruppenvariablen werden in die Kreuztabelle aufgenommen. Sie können wählen, in welcher Reihenfolge die Statistiken angezeigt werden. Außerdem werden Auswertungsstatistiken für jede Variable über alle Kategorien angezeigt. Die Datenwerte jeder Kategorie können aufgelistet oder unterdrückt werden. Bei umfangreichen Daten-Sets haben Sie die Möglichkeit, nur die ersten *n* Fälle aufzulisten.

Beispiel. Wie hoch liegen die durchschnittlichen Verkaufszahlen eines Produkts, gegliedert nach Region und Abnehmer? Möglicherweise stellen Sie fest, dass im Westen im Durchschnitt geringfügig mehr verkauft wird als in anderen Regionen, wobei gewerbliche Kunden in der westlichen Region die wichtigsten Abnehmer sind.

Statistiken. Summe, Anzahl der Fälle, Mittelwert, Median, gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozent der Gesamtsumme, Prozent der Gesamtanzahl (*N*), Prozent der Summe in, Prozent der Anzahl (*N*) in, geometrisches Mittel und harmonisches Mittel.

Daten. Die Gruppenvariablen stellen kategoriale Variablen dar, deren Werte numerisch oder Strings sein können. Die Anzahl der Kategorien sollte angemessen klein gehalten werden. Den anderen Variablen müssen Ränge zugeordnet werden können.

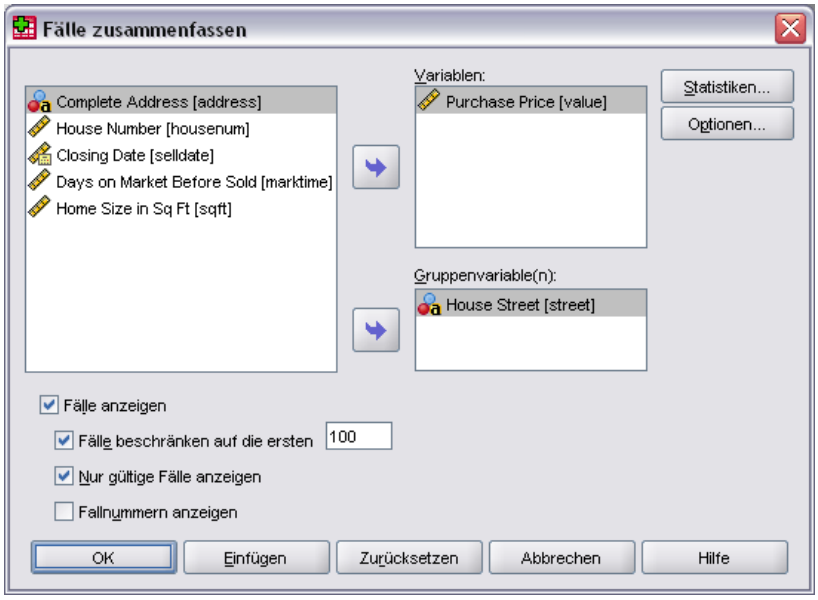
Annahmen. Einige der möglichen Untergruppenstatistiken, wie beispielsweise Mittelwert und Standardabweichung, basieren auf der Annahme, dass eine Normalverteilung vorliegt, und sind für Variablen mit symmetrischen Verteilungen geeignet. Robuste Statistiken, wie beispielsweise Median und Spannweite, sind für quantitative Variablen geeignet, die möglicherweise die Annahme einer Normalverteilung erfüllen.

So erstellen Sie Zusammenfassungen von Fällen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Berichte
    Fälle zusammenfassen

Abbildung 17-1
Dialogfeld "Fälle zusammenfassen"



▶ Wählen Sie mindestens eine Variable aus.

Die folgenden Optionen sind verfügbar:

- Sie können eine oder mehrere Gruppenvariablen auswählen, um die Daten in Untergruppen aufzuteilen.
- Klicken Sie auf Optionen, wenn Sie den Ausgabetitel ändern, eine Erklärung unter der Ausgabe hinzufügen oder Fälle mit fehlenden Werten ausschließen möchten.
- Sie können optionale Statistiken anzeigen lassen, indem Sie auf Statistik klicken.
- Wählen Sie Fälle anzeigen, um die Fälle in jeder Untergruppe auflisten zu lassen. In der Standardeinstellung werden nur die ersten 100 Fälle in der Datei aufgelistet. Sie können den Wert für Fälle beschränken auf die ersten*n* erhöhen oder vermindern bzw. diese Option deaktivieren, um alle Fälle auflisten zu lassen.

Zusammenfassen: Optionen

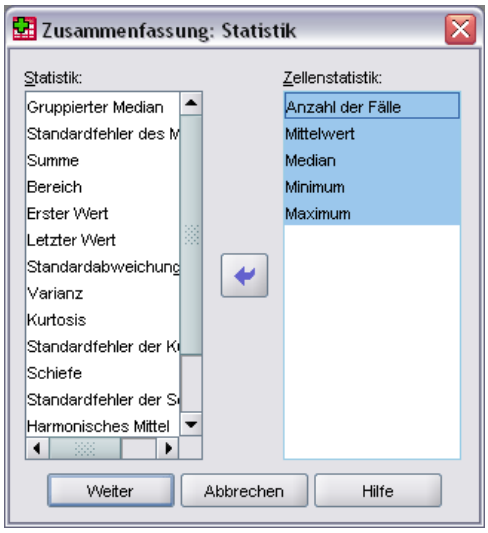
Abbildung 17-2
Dialogfeld "Optionen"

Sie können den Titel der Ausgabe ändern oder eine Erklärung hinzufügen, die unter der Ausgabetabelle angezeigt wird. Sie können den Zeilenumbruch in Titeln und Erklärungen steuern, indem Sie an die Stellen, an denen ein Zeilenumbruch durchgeführt werden soll, die Zeichen \n eingeben.

Außerdem können Sie Untertitel für Gesamtergebnisse ein- oder ausblenden sowie Fälle mit fehlenden Werten für beliebige, in der Analyse verwendete Variablen ein- oder ausschließen. Oft ist es angebracht, fehlende Fälle in der Ausgabe mit einem Punkt oder einem Sternchen zu kennzeichnen. Geben Sie ein Zeichen, eine Wortgruppe oder einen Code ein, der bei einem fehlenden Wert angezeigt werden soll, andernfalls werden fehlende Werte in der Ausgabe nicht besonders verarbeitet.

Zusammenfassung: Statistik

Abbildung 17-3
Dialogfeld "Zusammenfassung: Statistik"



Sie können mindestens eine der folgenden Untergruppen-Statistiken für die Variablen innerhalb jeder Kategorie jeder Gruppenvariablen auswählen: Summe, Anzahl der Fälle, Mittelwert, Median, gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozent der Gesamtsumme, Prozent der Gesamtanzahl, Prozent der Summe in, Prozent der Anzahl in, geometrisches Mittel und harmonisches Mittel. Die Statistiken werden in der Liste "Zellenstatistik" in derselben Reihenfolge angezeigt, in welcher sie in der Ausgabe angezeigt werden. Außerdem werden die Auswertungsstatistiken für jede Variable über alle Kategorien angezeigt.

Erster. Zeigt den ersten Datenwert in der Datendatei an.

Geometrisches Mittel. Die n-te Wurzel aus dem Produkt der Datenwerte, wobei n der Anzahl der Fälle entspricht.

Gruppierter Median. Der Median für Daten, die in Gruppen kodiert wurden (bei denen also ein Wert für ein ganzes Intervall steht). Wenn z. B. für das Alter jeder Wert in den Dreißigern als 35 kodiert ist, jeder Wert in den Vierzigern als 45 usw., dann wird der gruppierte Median aus den kodierten Daten berechnet.

Harmonisches Mittel. Wird verwendet, um die "mittlere" Gruppengröße zu bestimmen, wenn der Stichprobenumfang in den einzelnen Gruppen unterschiedlich ist. Das harmonische Mittel ist gleich der Gesamtzahl der Stichproben geteilt durch die Summe der Kehrwerte der Stichprobengrößen.

Kurtosis. Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flanken. Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flanken.

Letzter. Hiermit wird der letzte Datenwert in der Datendatei angezeigt.

Maximum. Der größte Wert einer numerischen Variablen.

Mittelwert. Ein Lagemaß. Das arithmetische Mittel, d. h. die Summe geteilt durch die Anzahl der Fälle.

Median. Wert, über und unter dem jeweils die Hälfte der Fälle liegt; 50. Perzentil. Bei einer geraden Anzahl von Fällen ist der Median der Mittelwert der beiden mittleren Fälle, wenn diese auf- oder absteigend sortiert sind. Der Median ist ein Lagemaß, das gegenüber Ausreißern unempfindlich ist (im Gegensatz zum Mittelwert, der durch wenige extrem niedrige oder hohe Werte beeinflusst werden kann).

Minimum. Der kleinste Wert einer numerischen Variablen.

N. Die Anzahl der Fälle (Beobachtungen oder Datensätze).

Prozent der Gesamtanzahl. Prozentsatz der Gesamtanzahl von Fällen in jeder Kategorie.

Prozent der Gesamtsumme. Prozentsatz der Gesamtsumme in jeder Kategorie.

Spannweite. Die Differenz zwischen den größten und kleinsten Werten einer numerischen Variablen; Maximalwert minus Minimalwert.

Schiefe. Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert 0. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (lange rechte Flanke). Eine Verteilung mit einer deutlichen negativen Schiefe läuft nach links lang aus (lange linke Flanke). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler, für eine Abweichung von der Symmetrie spricht.

Standardfehler der Kurtosis. Das Verhältnis der Kurtosis zu ihrem Standardfehler kann für einen Test auf Normalverteilung verwendet werden (d. h. die Annahme, dass Normalverteilung vorliegt, kann abgelehnt werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Kurtosis deutet darauf hin, dass die Flanken der Verteilung länger sind als bei einer Normalverteilung; ein negativer Wert bedeutet, dass sie kürzer sind (etwa wie bei einer kastenförmigen, gleichförmigen Verteilung).

Standardfehler der Schiefe. Das Verhältnis der Schiefe zu ihrem Standardfehler kann für einen Test auf Normalverteilung verwendet werden (d. h. die Annahme, dass Normalverteilung vorliegt, kann abgelehnt werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Schiefe bedeutet, dass die Verteilung eine lange rechte Flanke hat; ein extremer negativer Wert bedeutet, dass sie eine lange linke Flanke hat.

Summe. Die Summe der Werte über alle Fälle mit nichtfehlenden Werten.

Varianz. Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Maßeinheit der Varianz ist das Quadrat der Maßeinheiten der Variablen.

Kapitel
18

Mittelwerte

Mit der Prozedur “Mittelwerte” werden die Mittelwerte von Untergruppen und verwandte univariate Statistiken für abhängige Variablen innerhalb von Kategorien von mindestens einer unabhängigen Variablen berechnet. Wahlweise können Sie eine einfaktorielle Varianzanalyse, Eta und einen Test auf Linearität berechnen lassen.

Beispiel. Sie messen die mittlere Menge von Fett, die von drei verschiedenen Sorten Speiseöl absorbiert wird. Anschließend führen Sie eine einfaktorielle Varianzanalyse aus, um festzustellen, ob sich die Mittelwerte unterscheiden.

Statistiken. Summe, Anzahl der Fälle, Mittelwert, Median, gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozent der Gesamtsumme, Prozent der Gesamtanzahl (*N*), Prozent der Summe in, Prozent der Anzahl (*N*) in, geometrisches Mittel und harmonisches Mittel. Unter Optionen stehen außerdem Varianzanalyse, Eta, Eta-Quadrat, *R* und R^2 zur Verfügung.

Daten. Die abhängigen Variablen sind quantitativ, die unabhängigen Variablen kategorial. Die Werte der kategorialen Variablen können numerische Variablen oder String-Variablen sein.

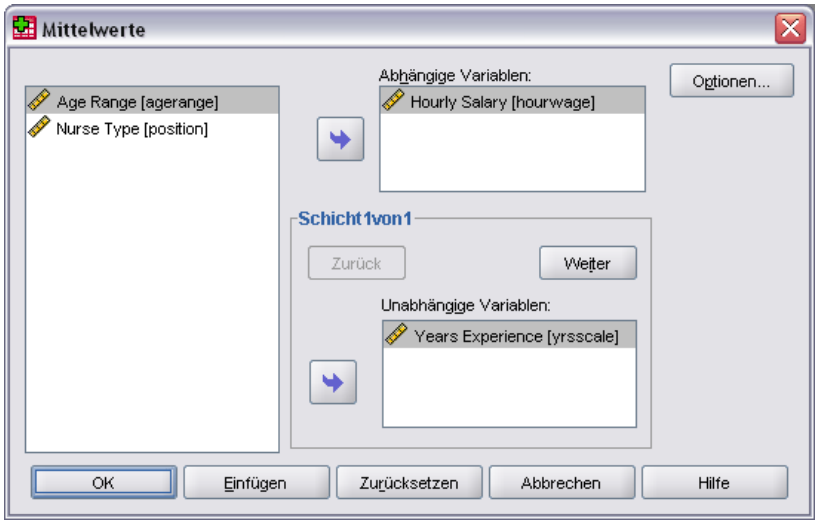
Annahmen. Einige der möglichen Untergruppenstatistiken, wie beispielsweise Mittelwert und Standardabweichung, basieren auf der Annahme, dass eine Normalverteilung vorliegt, und sind für Variablen mit symmetrischen Verteilungen geeignet. Robuste Statistiken, z. B. Median, sind für quantitative Variablen geeignet, die möglicherweise die Annahme einer Normalverteilung erfüllen. Die Varianzanalyse ist gegenüber Abweichungen von der Normalverteilung robust. Allerdings sollten die Daten in jeder Zelle symmetrisch sein. Bei der Varianzanalyse wird außerdem angenommen, dass die Gruppen aus Grundgesamtheiten mit gleichen Varianzen stammen. Zum Testen dieser Annahme können Sie den Levene-Test auf Homogenität der Varianzen verwenden. Dieser Test ist in der Prozedur “Einfaktorielle ANOVA” verfügbar.

So berechnen Sie die Mittelwerte der Untergruppen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mittelwerte vergleichen
    Mittelwerte

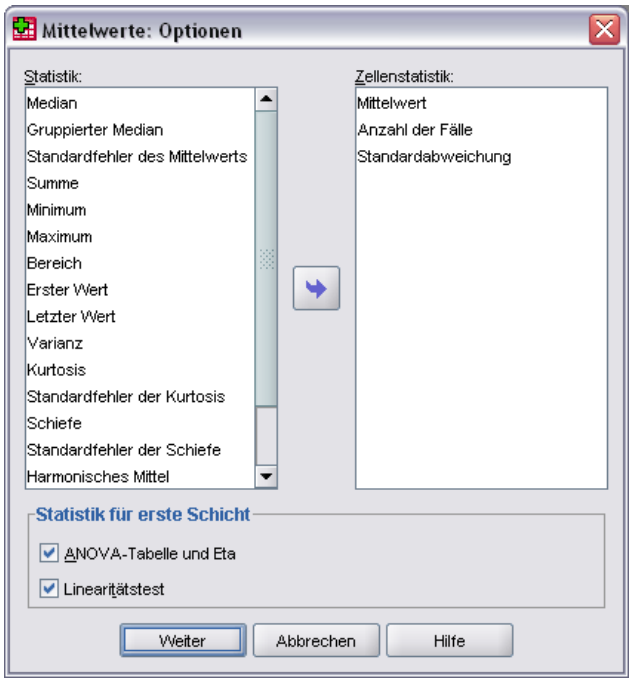
Abbildung 18-1
Dialogfeld "Mittelwerte"



- Wählen Sie eine oder mehrere abhängige Variablen aus.
- Verwenden Sie eine der folgenden Methoden, um die kategorialen unabhängigen Variablen auszuwählen:
  - Wählen Sie mindestens eine unabhängige Variable aus. Für jede unabhängige Variable werden getrennte Ergebnisse angezeigt.
  - Wählen Sie mindestens eine Schicht von unabhängigen Variablen aus. Die Stichprobe wird durch jede Schicht weiter unterteilt. Wenn es eine unabhängige Variable in Schicht 1 und eine unabhängige Variable in Schicht 2 gibt, werden die Ergebnisse nicht in einzelnen Tabellen für die unabhängigen Variablen, sondern in einer Kreuztabelle angezeigt.
- Sie können optionale Statistiken, eine Tabelle für die Varianzanalyse, Eta, Eta-Quadrat, R und R^2 berechnen lassen, indem Sie auf Optionen klicken.

Mittelwerte: Optionen

Abbildung 18-2
Dialogfeld "Mittelwerte: Optionen"



Sie können mindestens eine der folgenden Untergruppen-Statistiken für die Variablen innerhalb jeder Kategorie jeder Gruppenvariablen auswählen: Summe, Anzahl der Fälle, Mittelwert, Median, gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozent der Gesamtsumme, Prozent der Gesamtanzahl, Prozent der Summe in, Prozent der Anzahl in, geometrisches Mittel und harmonisches Mittel. Sie können die Reihenfolge ändern, in der die Statistiken für die Untergruppen berechnet werden. Die Statistiken werden in der Liste "Zellenstatistik" in derselben Reihenfolge angezeigt, in der sie in der Ausgabe angezeigt werden. Außerdem werden die Auswertungsstatistiken für jede Variable über alle Kategorien angezeigt.

Erster. Zeigt den ersten Datenwert in der Datendatei an.

Geometrisches Mittel. Die n-te Wurzel aus dem Produkt der Datenwerte, wobei n der Anzahl der Fälle entspricht.

Gruppierter Median. Der Median für Daten, die in Gruppen kodiert wurden (bei denen also ein Wert für ein ganzes Intervall steht). Wenn z. B. für das Alter jeder Wert in den Dreißigern als 35 kodiert ist, jeder Wert in den Vierzigern als 45 usw., dann wird der gruppierte Median aus den kodierten Daten berechnet.

Harmonisches Mittel. Wird verwendet, um die "mittlere" Gruppengröße zu bestimmen, wenn der Stichprobenumfang in den einzelnen Gruppen unterschiedlich ist. Das harmonische Mittel ist gleich der Gesamtzahl der Stichproben geteilt durch die Summe der Kehrwerte der Stichprobengrößen.

Kurtosis. Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flanken. Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flanken.

Letzter. Hiermit wird der letzte Datenwert in der Datendatei angezeigt.

Maximum. Der größte Wert einer numerischen Variablen.

Mittelwert. Ein Lagemaß. Das arithmetische Mittel, d. h. die Summe geteilt durch die Anzahl der Fälle.

Median. Wert, über und unter dem jeweils die Hälfte der Fälle liegt; 50. Perzentil. Bei einer geraden Anzahl von Fällen ist der Median der Mittelwert der beiden mittleren Fälle, wenn diese auf- oder absteigend sortiert sind. Der Median ist ein Lagemaß, das gegenüber Ausreißern unempfindlich ist (im Gegensatz zum Mittelwert, der durch wenige extrem niedrige oder hohe Werte beeinflusst werden kann).

Minimum. Der kleinste Wert einer numerischen Variablen.

N. Die Anzahl der Fälle (Beobachtungen oder Datensätze).

Prozent der Gesamtanzahl. Prozentsatz der Gesamtanzahl von Fällen in jeder Kategorie.

Prozent der Gesamtsumme. Prozentsatz der Gesamtsumme in jeder Kategorie.

Spannweite. Die Differenz zwischen den größten und kleinsten Werten einer numerischen Variablen; Maximalwert minus Minimalwert.

Schiefe. Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert 0. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (lange rechte Flanke). Eine Verteilung mit einer deutlichen negativen Schiefe läuft nach links lang aus (lange linke Flanke). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler, für eine Abweichung von der Symmetrie spricht.

Standardfehler der Kurtosis. Das Verhältnis der Kurtosis zu ihrem Standardfehler kann für einen Test auf Normalverteilung verwendet werden (d. h. die Annahme, dass Normalverteilung vorliegt, kann abgelehnt werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Kurtosis deutet darauf hin, dass die Flanken der Verteilung länger sind als bei einer Normalverteilung; ein negativer Wert bedeutet, dass sie kürzer sind (etwa wie bei einer kastenförmigen, gleichförmigen Verteilung).

Standardfehler der Schiefe. Das Verhältnis der Schiefe zu ihrem Standardfehler kann für einen Test auf Normalverteilung verwendet werden (d. h. die Annahme, dass Normalverteilung vorliegt, kann abgelehnt werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Schiefe bedeutet, dass die Verteilung eine lange rechte Flanke hat; ein extremer negativer Wert bedeutet, dass sie eine lange linke Flanke hat.

Summe. Die Summe der Werte über alle Fälle mit nichtfehlenden Werten.

Varianz. Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Maßeinheit der Varianz ist das Quadrat der Maßeinheiten der Variablen.

Statistik für erste Schicht

ANOVA-Tabelle und Eta. Zeigt eine Tabelle für eine einfaktorielle Varianzanalyse an und berechnet Eta und Eta-Quadrat (Zusammenhangsmaße) für jede unabhängige Variable in der ersten Schicht.

Linearitätstest. Berechnet für lineare und nichtlineare Komponenten die Quadratsummen, die Freiheitsgrade und das Mittel der Quadrate sowie den F-Wert, R und R-Quadrat. Die Berechnungen für Linearität werden nicht durchgeführt, wenn die unabhängige Variable eine kurze String-Variable ist.

Kapitel
19

OLAP-Würfel

Mit der Prozedur “OLAP-Würfel” (Online Analytical Processing) werden Gesamtwerte, Mittelwerte und andere univariate Statistiken für stetige Auswertungsvariablen innerhalb der Kategorien von mindestens einer kategorialen Gruppenvariablen berechnet. Für jede Kategorie der Gruppenvariablen wird eine separate Schicht erstellt.

Beispiel. Durchschnittlicher und gesamter Umsatz für verschiedene Regionen und Produktlinien innerhalb einer Region.

Statistiken. Summe, Anzahl der Fälle, Mittelwert, Median, Gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozentsatz der gesamten Fälle, Prozentsatz der Gesamtsumme, Prozentsatz der gesamten Fälle innerhalb der Gruppenvariablen, Prozentsatz der Gesamtsumme innerhalb der Gruppenvariablen, geometrisches Mittel und harmonisches Mittel.

Daten. Die Auswertungsvariablen sind quantitativ (stetige Variablen, die auf einer Intervall- oder Verhältnisskala gemessen werden) und die Gruppenvariablen kategorial. Die Werte der kategorialen Variablen können numerische Variablen oder String-Variablen sein.

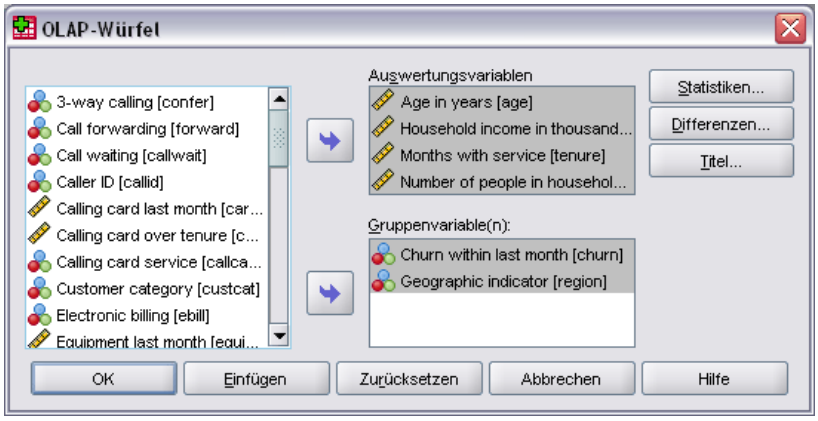
Annahmen. Einige der möglichen Untergruppenstatistiken, wie beispielsweise Mittelwert und Standardabweichung, basieren auf der Annahme, dass eine Normalverteilung vorliegt, und sind für Variablen mit symmetrischen Verteilungen geeignet. Robuste Statistiken, wie z. B. Median und Spannweite, sind für quantitative Variablen geeignet, die möglicherweise die Annahme einer Normalverteilung erfüllen.

So erstellen Sie OLAP-Würfel:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Berichte
    OLAP-Würfel...

Abbildung 19-1
Dialogfeld "OLAP-Würfel"



- ▶ Wählen Sie mindestens eine stetige Auswertungsvariable aus.
- ▶ Wählen Sie mindestens eine kategoriale Gruppenvariable aus.

Die folgenden Optionen sind verfügbar:

- Sie können verschiedene Auswertungsstatistiken auswählen, indem Sie auf Statistiken klicken. Sie müssen mindestens eine Gruppenvariable auswählen, bevor Sie die Auswertungsstatistiken auswählen können.
- Sie können die Differenzen zwischen Variablenpaaren und Gruppenpaaren berechnen lassen, die durch eine Gruppenvariable definiert sind, indem Sie auf Differenzen klicken.
- Sie können Titel für benutzerdefinierte Tabellen erstellen, indem Sie auf Titel klicken.

OLAP-Würfel: Statistiken

Abbildung 19-2
Dialogfeld "OLAP-Würfel: Statistiken"

Sie können eine oder mehrere der folgenden Untergruppen-Statistiken für die Auswertungsvariablen in jeder Kategorie aller Gruppenvariablen auswählen: Summe, Anzahl der Fälle, Mittelwert, Median, Gruppierter Median, Standardfehler des Mittelwerts, Minimum, Maximum, Spannweite, Variablenwert der ersten Kategorie der Gruppenvariablen, Variablenwert der letzten Kategorie der Gruppenvariablen, Standardabweichung, Varianz, Kurtosis, Standardfehler der Kurtosis, Schiefe, Standardfehler der Schiefe, Prozentsatz der gesamten Fälle, Prozentsatz der Gesamtsumme, Prozentsatz der gesamten Fälle innerhalb der Gruppenvariablen, Prozentsatz der Gesamtsumme innerhalb der Gruppenvariablen, geometrisches Mittel und harmonisches Mittel.

Sie können die Reihenfolge ändern, in der die Statistiken für die Untergruppen berechnet werden. Die Statistiken werden in der Liste "Zellenstatistik" in derselben Reihenfolge angezeigt, in der sie in der Ausgabe angezeigt werden. Außerdem werden die Auswertungsstatistiken für jede Variable über alle Kategorien angezeigt.

Erster. Zeigt den ersten Datenwert in der Datendatei an.

Geometrisches Mittel. Die n-te Wurzel aus dem Produkt der Datenwerte, wobei n der Anzahl der Fälle entspricht.

Gruppierter Median. Der Median für Daten, die in Gruppen kodiert wurden (bei denen also ein Wert für ein ganzes Intervall steht). Wenn z. B. für das Alter jeder Wert in den Dreißigern als 35 kodiert ist, jeder Wert in den Vierzigern als 45 usw., dann wird der gruppierte Median aus den kodierten Daten berechnet.

Harmonisches Mittel. Wird verwendet, um die "mittlere" Gruppengröße zu bestimmen, wenn der Stichprobenumfang in den einzelnen Gruppen unterschiedlich ist. Das harmonische Mittel ist gleich der Gesamtzahl der Stichproben geteilt durch die Summe der Kehrwerte der Stichprobengrößen.

Kurtosis. Ein Maß dafür, wie sich die Beobachtungen um einen zentralen Punkt gruppieren. Bei einer Normalverteilung ist der Wert der Kurtosis gleich 0. Bei positiver Kurtosis gruppieren sich die Beobachtungen dichter als bei der Normalverteilung und haben längere Flanken. Bei negativer Kurtosis gruppieren sich die Beobachtungen weniger dicht zusammen und haben kürzere Flanken.

Letzter. Hiermit wird der letzte Datenwert in der Datendatei angezeigt.

Maximum. Der größte Wert einer numerischen Variablen.

Mittelwert. Ein Lagemaß. Das arithmetische Mittel, d. h. die Summe geteilt durch die Anzahl der Fälle.

Median. Wert, über und unter dem jeweils die Hälfte der Fälle liegt; 50. Perzentil. Bei einer geraden Anzahl von Fällen ist der Median der Mittelwert der beiden mittleren Fälle, wenn diese auf- oder absteigend sortiert sind. Der Median ist ein Lagemaß, das gegenüber Ausreißern unempfindlich ist (im Gegensatz zum Mittelwert, der durch wenige extrem niedrige oder hohe Werte beeinflusst werden kann).

Minimum. Der kleinste Wert einer numerischen Variablen.

N. Die Anzahl der Fälle (Beobachtungen oder Datensätze).

Prozent der Anzahl in. Prozentsatz der Gesamtanzahl von Fällen für die angegebene Gruppenvariable in den Kategorien der anderen Gruppenvariablen. Wenn nur eine Gruppenvariable vorhanden ist, ist dieser Wert gleich dem Prozentsatz der Gesamtanzahl von Fällen.

Prozent der Summe in. Prozentsatz der Summe für die angegebene Gruppenvariable in den Kategorien der anderen Gruppenvariablen. Wenn nur eine Gruppenvariable vorhanden ist, ist dieser Wert gleich dem Prozentsatz der Gesamtsumme.

Prozent der Gesamtanzahl. Prozentsatz der Gesamtanzahl von Fällen in jeder Kategorie.

Prozent der Gesamtsumme. Prozentsatz der Gesamtsumme in jeder Kategorie.

Spannweite. Die Differenz zwischen den größten und kleinsten Werten einer numerischen Variablen; Maximalwert minus Minimalwert.

Schiefe. Ein Maß für die Asymmetrie einer Verteilung. Die Normalverteilung ist symmetrisch, ihre Schiefe hat den Wert 0. Eine Verteilung mit einer deutlichen positiven Schiefe läuft nach rechts lang aus (lange rechte Flanke). Eine Verteilung mit einer deutlichen negativen Schiefe läuft nach links lang aus (lange linke Flanke). Als Faustregel kann man verwenden, dass ein Schiefe-Wert, der mehr als doppelt so groß ist wie sein Standardfehler, für eine Abweichung von der Symmetrie spricht.

Standardfehler der Kurtosis. Das Verhältnis der Kurtosis zu ihrem Standardfehler kann für einen Test auf Normalverteilung verwendet werden (d. h. die Annahme, dass Normalverteilung vorliegt, kann abgelehnt werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer positiver Wert für die Kurtosis deutet darauf hin, dass die Flanken der Verteilung länger sind als bei einer Normalverteilung; ein negativer Wert bedeutet, dass sie kürzer sind (etwa wie bei einer kastenförmigen, gleichförmigen Verteilung).

Standardfehler der Schiefe. Das Verhältnis der Schiefe zu ihrem Standardfehler kann für einen Test auf Normalverteilung verwendet werden (d. h. die Annahme, dass Normalverteilung vorliegt, kann abgelehnt werden, wenn das Verhältnis kleiner als -2 oder größer als +2 ist). Ein großer

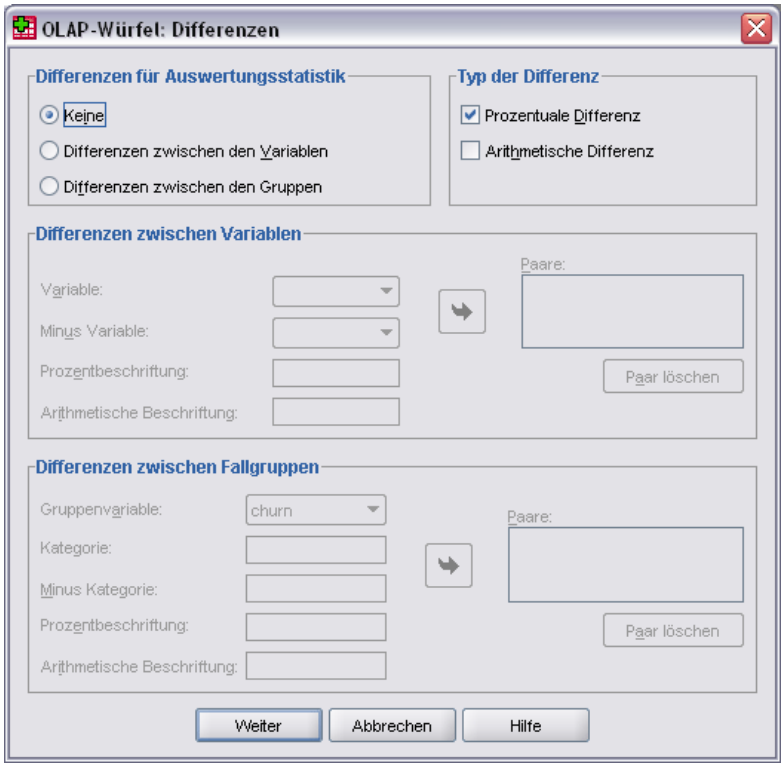
positiver Wert für die Schiefe bedeutet, dass die Verteilung eine lange rechte Flanke hat; ein extremer negativer Wert bedeutet, dass sie eine lange linke Flanke hat.

Summe. Die Summe der Werte über alle Fälle mit nichtfehlenden Werten.

Varianz. Ein Maß der Streuung um den Mittelwert. Es ist gleich dem Quotienten aus der Summe der quadrierten Abweichung vom Mittelwert und der um 1 verringerten Fallanzahl. Die Maßeinheit der Varianz ist das Quadrat der Maßeinheiten der Variablen.

OLAP-Würfel: Differenzen

Abbildung 19-3
Dialogfeld "OLAP-Würfel: Differenzen"



In diesem Dialogfeld können Sie prozentuale und arithmetische Differenzen zwischen Auswertungsvariablen oder zwischen Gruppen berechnen lassen, die durch eine Gruppenvariable definiert sind. Die Differenzen werden für alle Maße berechnet, die im Dialogfeld "OLAP-Würfel: Statistiken" ausgewählt wurden.

Differenzen zwischen den Variablen. Hiermit werden die Differenzen zwischen Variablenpaaren berechnet. Die Werte der Auswertungsstatistik für die zweite Variable (die Minusvariable) in jedem Paar werden von den Werten der Auswertungsstatistik für die erste Variable im Paar subtrahiert. Bei prozentualen Differenzen wird der Wert der Auswertungsvariable für die Minusvariable als Nenner verwendet. Sie müssen mindestens zwei Auswertungsvariablen im Hauptdialogfeld auswählen, bevor Sie die Differenzen zwischen den Variablen angeben können.

Differenzen zwischen Fallgruppen. Hiermit werden die Differenzen zwischen Gruppenpaaren berechnet, die durch eine Gruppenvariable definiert sind. Die Werte der Auswertungsstatistik für die zweite Kategorie (die Minuskategorie) in jedem Paar werden von den Werten der Auswertungsstatistik für die erste Kategorie im Paar subtrahiert. Bei prozentualen Differenzen wird der Wert der Auswertungsstatistik für die Minuskategorie als Nenner verwendet. Sie müssen mindestens eine Gruppenvariable im Hauptdialogfeld auswählen, bevor Sie die Differenzen zwischen den Gruppen angeben können.

OLAP-Würfel: Titel

Abbildung 19-4
Dialogfeld "OLAP-Würfel: Titel"

Sie können den Titel der Ausgabe ändern oder eine Erklärung hinzufügen, die unter der Ausgabetabelle angezeigt wird. Sie können auch den Zeilenumbruch in Titeln und Erklärungen selbst bestimmen, indem Sie an der gewünschten Stelle im Text die Zeichenfolge \n eingeben.

Kapitel
20

T-Tests

Es sind drei Typen von *T*-Tests verfügbar:

T-Test bei unabhängigen Stichproben (T-Test bei zwei Stichproben). Vergleicht die Mittelwerte einer Variablen für zwei Fallgruppen. Für jede Gruppe sind beschreibende Statistiken und der Levene-Test auf Gleichheit der Varianzen sowie *t*-Werte für gleiche und verschiedene Varianzen und ein 95%-Konfidenzintervall für die Differenz der Mittelwerte verfügbar.

T-Test bei gepaarten Stichproben (T-Test für abhängige Variablen). Vergleicht den Mittelwert von zwei Variablen für eine einzelne Gruppe. Dieser Test ist auch für Studien mit zugeordneten Paaren oder Fallkontrolle geeignet. Die Ausgabe enthält deskriptive Statistiken für die Testvariablen, die Korrelationen zwischen den Variablen, deskriptive Statistiken für die gepaarten Differenzen, den *T*-Test und ein 95%-Konfidenzintervall.

T-Test bei einer Stichprobe. Vergleicht den Mittelwert einer Variablen mit einem bekannten oder angenommenen Wert. Neben dem *T*-Test werden deskriptive Statistiken für die Testvariablen angezeigt. In der Standardeinstellung wird unter anderem ein 95%-Konfidenzintervall für die Differenz zwischen dem Mittelwert der Testvariablen und dem angenommenen Testwert ausgegeben.

T-Test bei unabhängigen Stichproben

Im T-Test bei unabhängigen Stichproben werden die Mittelwerte von zwei Fallgruppen verglichen. Im Idealfall sollten die Subjekte bei diesem Test zufällig zwei Gruppen zugeordnet werden, sodass Unterschiede bei den Antworten lediglich auf die Behandlung (bzw. Nichtbehandlung) und keine sonstigen Faktoren zurückzuführen sind. Dies ist nicht der Fall, wenn Sie die Durchschnittseinkommen von Männern und Frauen vergleichen. Die jeweiligen Personen sind nicht zufällig auf die Gruppen "männlich" oder "weiblich" verteilt. In solchen Situationen müssen Sie sicherstellen, dass signifikante Differenzen der Mittelwerte nicht durch Abweichungen bei anderen Faktoren verborgen oder verstärkt werden. Unterschiede im Durchschnittseinkommen können auch durch Faktoren wie den Bildungsstand beeinflußt werden (nicht nur durch das Geschlecht).

Beispiel. Patienten mit hohem Blutdruck werden zufällig auf eine Kontrollgruppe und eine Versuchsgruppe verteilt. Die Patienten in der Kontrollgruppe erhalten ein Plazebo. Die Patienten der Versuchsgruppe erhalten ein neues Medikament, dessen blutdrucksenkende Wirkung erprobt werden soll. Nach zweimonatiger Behandlung wird der *T*-Test bei zwei Stichproben angewandt, um den durchschnittlichen Blutdruck der Personen in der Kontrollgruppe mit dem der Personen aus der Versuchsgruppe zu vergleichen. Bei jedem Patienten wird eine Messung vorgenommen, und er gehört zu jeweils einer (1) Gruppe.

Statistiken. Für jede Variable: Stichprobengröße, Mittelwert, Standardabweichung und Standardfehler des Mittelwerts. Für die Differenz der Mittelwerte: Mittelwert, Standardfehler und Konfidenzintervall. (Sie können das Konfidenzniveau bestimmen.) Tests: Levene-Test auf Gleichheit der Varianzen sowie *t*-Tests auf Gleichheit der Mittelwerte bei gemeinsamen und separaten Varianzen.

Daten. Die Werte der untersuchten quantitativen Variablen müssen in einer einzelnen Spalte in der Datendatei vorliegen. Zum Aufteilen der Fälle in zwei Gruppen verwendet SPSS eine Gruppenvariable mit zwei Werten. Die Gruppenvariable kann numerische Werte (wie zum Beispiel 1 und 2 oder 6,25 und 12,5) oder kurze Strings (beispielsweise *Ja* und *Nein*) enthalten. Alternativ können Sie eine quantitative Variable wie z. B. *Alter* verwenden und die Fälle durch Angabe eines Trennwerts aufteilen (der Trennwert 21 teilt *Alter* in eine Gruppe "unter 21" und eine "21 und darüber").

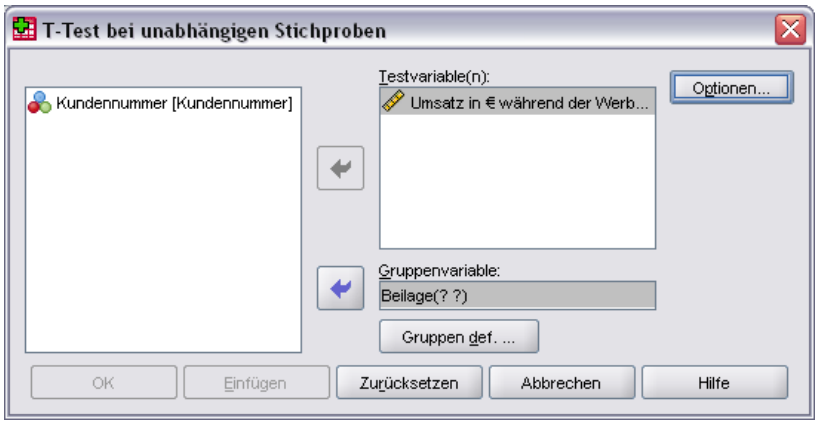
Annahmen. Für den *T*-Test auf Gleichheit der Varianzen sollten die Beobachtungen unabhängige Zufallsstichproben aus Normalverteilungen mit derselben Varianz der Grundgesamtheit sein. Für den *T*-Test auf Ungleichheit der Varianzen sollten die Beobachtungen unabhängige Zufallsstichproben aus Normalverteilungen sein. Der *T*-Test mit zwei Stichproben ist relativ robust gegenüber Abweichungen von der Normalverteilung. Achten Sie bei der grafischen Überprüfung von Verteilungen darauf, dass diese symmetrisch sind und keine Ausreißer enthalten.

So lassen Sie einen T-Test bei unabhängigen Stichproben berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mittelwerte vergleichen
    T-Test bei unabhängigen Stichproben...

Abbildung 20-1
Dialogfeld "T-Test bei unabhängigen Stichproben"



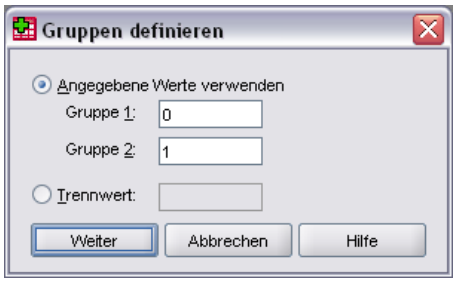
▶ Wählen Sie mindestens eine quantitative Testvariable. Für jede Variable wird ein separater *T*-Test berechnet.

▶ Wählen Sie eine einzelne Gruppenvariable aus und klicken Sie dann auf Gruppen def., um zwei Codes für die zu vergleichenden Gruppen anzugeben.

- Zusätzlich können Sie auf Optionen klicken, um die Behandlung fehlender Daten und das Niveau des Konfidenzintervalls festzulegen.

T-Test bei unabhängigen Stichproben: Gruppen definieren

Abbildung 20-2
Dialogfeld "Gruppen definieren" für numerische Variablen



Definieren Sie bei numerischen Gruppenvariablen die zwei Gruppen für den *t*-Test, indem Sie zwei Werte oder einen Trennwert angeben:

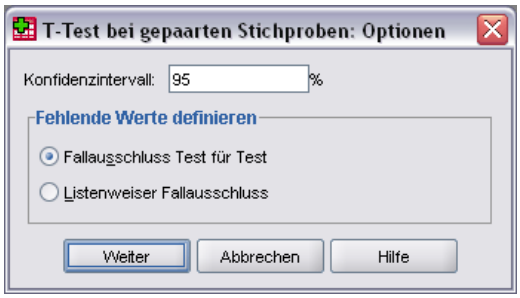
- **Angegebene Werte verwenden.** Geben Sie einen Wert für Gruppe 1 und einen weiteren Wert für Gruppe 2 ein. Fälle mit anderen Werten werden aus der Analyse ausgeschlossen. Zahlen müssen nichtganzzahlig sein (so sind beispielsweise 6,25 und 12,5 gültige Werte).
- **Trennwert.** Geben Sie eine Zahl ein, welche die Werte der Gruppenvariablen in zwei Mengen aufteilt. Alle Fälle mit Werten, die kleiner als der Trennwert sind, bilden eine Gruppe. Die Fälle mit Werten größer oder gleich dem Trennwert bilden die andere Gruppe.

Abbildung 20-3
Dialogfeld "Gruppen definieren" für String-Variablen

Bei String-Gruppenvariablen geben Sie einen String für Gruppe 1 und einen anderen für Gruppe 2 ein, beispielsweise *Ja* und *Nein*. Fälle mit anderen Strings werden von der Analyse ausgeschlossen.

T-Tests bei unabhängigen Stichproben: Optionen

Abbildung 20-4
Dialogfeld "T-Test bei unabhängigen Stichproben: Optionen"



Konfidenzintervall. In der Standardeinstellung wird ein 95%-Konfidenzintervall für die Differenz der Mittelwerte angezeigt. Geben Sie einen Wert zwischen 1 und 99 ein, um ein anderes Konfidenzniveau festzulegen.

Fehlende Werte. Wenn Sie mehrere Variablen testen und bei einer oder mehreren Variablen Daten fehlen, können Sie bestimmen, welche Fälle einzuschließen (oder auszuschließen) sind.

- **Fallausschluss Test für Test.** Bei jedem *T*-Test werden alle Fälle verwendet, für die gültige Daten für die getestete Variable vorliegen. Die Stichprobengröße kann von Test zu Test unterschiedlich ausfallen.
- **Listenweiser Fallausschluss.** Jeder *T*-Test verwendet nur Fälle mit gültigen Daten für alle in den angeforderten *T*-Tests verwendeten Variablen. Die Stichprobengröße bleibt bei allen Tests konstant.

T-Test bei gepaarten Stichproben

Mit der Prozedur "T-Test bei gepaarten Stichproben" werden die Mittelwerte zweier Variablen für eine einzelne Gruppe verglichen. Diese Prozedur berechnet für jeden Fall die Differenzen zwischen den Werten der zwei Variablen und überprüft, ob der Durchschnitt von 0 abweicht.

Beispiel. In einer Studie über Bluthochdruck wird der Blutdruck aller Patienten zu Beginn der Studie und nach der Behandlung gemessen. Daher gibt es für jede Testperson zwei Messwerte, die auch als *Vorher*- und *Nachher*-Messung bezeichnet werden. Dieser Test kann auch bei Studien mit zugeordneten Paaren bzw. mit Fallkontrolle verwendet werden. Hierbei enthält jeder Datensatz der Datendatei die Reaktion des Patienten und die von der zugehörigen Kontroll-Testperson. In einer Blutdruckstudie könnten den Patienten die Kontrollpersonen nach Alter zugeordnet werden (einem 75-jährigen Patienten ein 75-jähriges Mitglied der Kontrollgruppe).

Statistiken. Für jede Variable: Mittelwert, Stichprobengröße, Standardabweichung und Standardfehler des Mittelwerts. Für jedes Variablenpaar: Korrelation, durchschnittliche Differenz der Mittelwerte, *T*-Test und Konfidenzintervall für die Differenz der Mittelwerte. (Sie können das Konfidenzniveau festlegen.) Standardabweichung und Standardfehler der Differenz der Mittelwerte.

Daten. Legen Sie für jeden gepaarten Test zwei Variablen fest, die auf Intervall-Messniveau oder Verhältnis-Messniveau quantitativ sein müssen. In einer Studie mit zugeordneten Paaren bzw. mit Fallkontrolle müssen die Reaktionen jedes Testsubjektes und dessen zugeordneten Kontrollsubjektes im selben Fall der Datendatei enthalten sein.

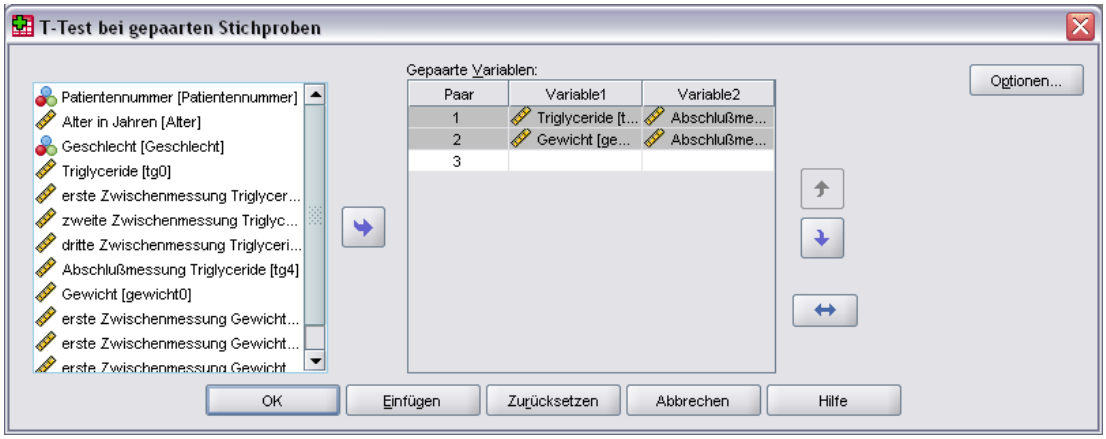
Annahmen. Die Beobachtungen für jedes Paar müssen unter gleichen Bedingungen vorgenommen werden. Die Differenzen der Mittelwerte müssen normalverteilt sein. Die Varianzen jeder Variablen können gleich oder ungleich sein.

So lassen Sie einen T-Test bei gepaarten Stichproben berechnen:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Mittelwerte vergleichen
  T-Test bei gepaarten Stichproben

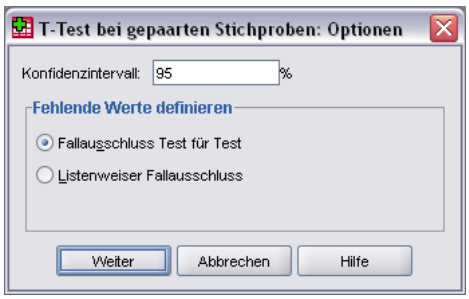
Abbildung 20-5
Dialogfeld "T-Test bei gepaarten Stichproben"



- Wählen Sie mindestens ein Variablenpaar aus
- Zusätzlich können Sie auf Optionen klicken, um die Behandlung fehlender Daten und das Niveau des Konfidenzintervalls festzulegen.

T-Test bei gepaarten Stichproben: Optionen

Abbildung 20-6
Dialogfeld "T-Test bei gepaarten Stichproben: Optionen"



Konfidenzintervall. In der Standardeinstellung wird ein 95%-Konfidenzintervall für die Differenz der Mittelwerte angezeigt. Geben Sie einen Wert zwischen 1 und 99 ein, um ein anderes Konfidenzniveau festzulegen.

Fehlende Werte. Wenn Sie mehrere Variablen testen und bei einer oder mehreren Variablen Daten fehlen, können Sie bestimmen, welche Fälle einzuschließen (oder auszuschließen) sind:

- **Fallausschluss Test für Test.** Bei jedem *T*-Test werden alle Fälle mit gültigen Daten für die getesteten Variablenpaare verwendet. Die Stichprobengröße kann von Test zu Test unterschiedlich ausfallen.
- **Listenweiser Fallausschluss.** Bei jedem *T*-Test werden nur Fälle mit gültigen Daten für alle getesteten Variablenpaare verwendet. Die Stichprobengröße bleibt bei allen Tests konstant.

T-Test bei einer Stichprobe

Die Prozedur "T-Test bei einer Stichprobe" prüft, ob der Mittelwert einer einzelnen Variablen von einer angegebenen Konstanten abweicht.

Beispiele. Ein Forscher könnte testen, ob der durchschnittliche IQ-Wert einer Gruppe von Studenten von 100 abweicht. Ein Hersteller von Getreideprodukten könnte stichprobenartig Packungen aus der Produktion entnehmen und prüfen, ob das Durchschnittsgewicht der Stichproben auf dem 95%-Konfidenzniveau von 500 Gramm abweicht.

Statistiken. Für jede Testvariable: Mittelwert, Standardabweichung und Standardfehler der Differenz der Mittelwerte. Außerdem die durchschnittliche Differenz zwischen jedem Datenwert und dem angenommenen Testwert, ein *T*-Test, der prüft, ob diese Differenz null beträgt, und ein Konfidenzintervall für diese Differenz. (Sie können das Konfidenzniveau festlegen.)

Daten. Um die Werte einer quantitativen Variablen mit einem angenommenen Testwert zu vergleichen, wählen Sie eine quantitative Variable aus und geben Sie einen angenommenen Testwert ein.

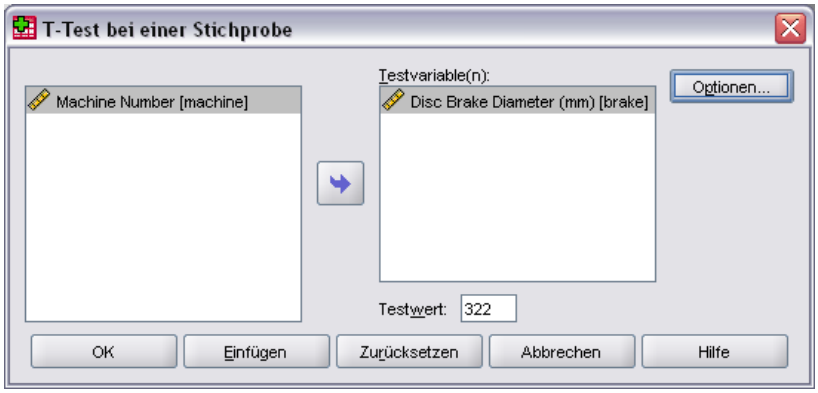
Annahmen. Bei diesem Test wird von einer Normalverteilung ausgegangen; er ist jedoch recht robust gegenüber Abweichungen von dieser Verteilung.

So lassen Sie den T-Test bei einer Stichprobe berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
Mittelwerte vergleichen
T-Test bei einer Stichprobe…

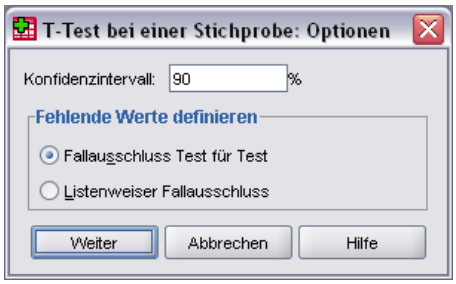
Abbildung 20-7
Dialogfeld "T-Test bei einer Stichprobe"



- Wählen Sie eine oder mehrere Variablen aus, die mit demselben angenommenen Wert verglichen werden sollen.
- Geben Sie einen numerischen Testwert ein, mit dem jeder Stichprobenmittelwert verglichen werden soll.
- Zusätzlich können Sie auf Optionen klicken, um die Behandlung fehlender Daten und das Niveau des Konfidenzintervalls festzulegen.

T-Test bei einer Stichprobe: Optionen

Abbildung 20-8
Dialogfeld "T-Test bei einer Stichprobe: Optionen"



Konfidenzintervall. In der Standardeinstellung wird ein 95%-Konfidenzintervall für die Differenz zwischen dem Mittelwert und dem angenommenen Testwert angezeigt. Geben Sie einen Wert zwischen 1 und 99 ein, um ein anderes Konfidenzniveau festzulegen.

Fehlende Werte. Wenn Sie mehrere Variablen testen und bei einer oder mehreren Variablen Daten fehlen, können Sie bestimmen, welche Fälle einzuschließen (oder auszuschließen) sind.

- **Fallausschluss Test für Test.** Bei jedem *T*-Test werden alle Fälle verwendet, die gültige Daten für die getestete Variable aufweisen. Die Stichprobengröße kann von Test zu Test unterschiedlich ausfallen.
- **Listenweiser Fallausschluss.** Jeder *T*-Test verwendet nur Fälle, die gültige Daten für alle Variablen aufweisen, die in einem der angeforderten *T*-Tests verwendet werden. Die Stichprobengröße bleibt bei allen Tests konstant.

Zusätzliche Funktionen beim Befehl T-TEST

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Erstellen von T-Tests für eine Stichprobe sowie für unabhängige Stichproben mit einem einzigen Befehl.
- Testen einer Variablen gegen alle Variablen in einer Liste mit einem gepaarten t-Test (mit dem Unterbefehl `PAIRS`).

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
21

Einfaktorielle ANOVA

Die Prozedur Einfaktorielle ANOVA führt eine einfaktorielle Varianzanalyse für eine quantitative abhängige Variable mit einer einzelnen (unabhängigen) Faktorvariablen durch. Mit der Varianzanalyse wird die Hypothese überprüft, dass mehrere Mittelwerte gleich sind. Dieses Verfahren ist eine Erweiterung des *T*-Tests bei zwei Stichproben.

Sie können zusätzlich zur Feststellung, dass Differenzen zwischen Mittelwerten vorhanden sind, auch bestimmen, welche Mittelwerte abweichen. Für den Vergleich von Mittelwerten gibt es zwei Arten von Tests: A-priori-Kontraste und Post-Hoc-Tests. Kontraste sind Tests, die *vor* der Ausführung des Experiments eingerichtet werden, Post-Hoc-Tests werden *nach* dem Experiment ausgeführt. Sie können auch auf Trends für mehrere Kategorien testen.

Beispiel. Paniertes Fleisch absorbiert beim Fritieren unterschiedliche Mengen an Fett. Ein Experiment wird mit den folgenden drei Fettsorten durchgeführt: Distelöl, Maiskeimöl und Schmalz. Distelöl und Maiskeimöl sind ungesättigte Fette, Schmalz ist ein gesättigtes Fett. Sie können bestimmen, ob die Menge des absorbierten Fetts von der Fettsorte abhängt. Gleichzeitig können Sie einen A-priori-Kontrast einrichten, um zu ermitteln, ob sich die absorbierte Fettmenge bei gesättigten und ungesättigten Fetten unterscheidet.

Statistiken. Für jede Gruppe: Anzahl der Fälle, Mittelwert, Standardabweichung, Standardfehler des Mittelwerts, Minimum, Maximum und 95%-Konfidenzintervall für den Mittelwert. Levene-Test auf Homogenität der Varianzen, Varianzanalyse-Tabellen und zuverlässige Tests auf Gleichheit der Mittelwerte für jede abhängige Variable, benutzerspezifische A-priori-Kontraste, Post-Hoc-Spannweitentests und Mehrfachvergleiche: Bonferroni, Sidak, ehrlich signifikante Differenz nach Tukey, GT2 nach Hochberg, Gabriel, *F*-Test nach Dunnett, Ryan-Einot-Gabriel-Welsch (*F* nach R-E-G-W), Spannweitentest nach Ryan-Einot-Gabriel-Welsch (*Q* nach R-E-G-W), Tamhane-T2, Dunnett-T3, Games-Howell, Dunnett-*C*, Duncans multipler Spannweitentest, Student-Newman-Keuls (S-N-K), Tukey-*b*, Waller-Duncan, Scheffé und geringste signifikante Differenz.

Daten. Die Werte der Faktorvariablen müssen ganzzahlig sein, die abhängige Variable muss quantitativ sein (Messung auf Intervallebene).

Annahmen. Jede Gruppe bildet eine unabhängige zufällige Stichprobe aus einer normalverteilten Grundgesamtheit. Die Varianzanalyse ist unempfindlich gegenüber Abweichungen von der Normalverteilung. Die Daten müssen jedoch symmetrisch verteilt sein. Die Gruppen müssen aus Grundgesamtheiten mit gleichen Varianzen stammen. Sie überprüfen diese Annahme mithilfe des Levene-Tests auf Homogenität der Varianzen.

So lassen Sie eine einfaktorielle ANOVA berechnen:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

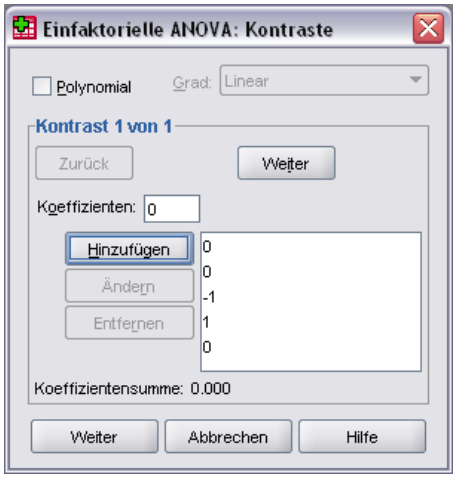
  Analysieren
  Mittelwerte vergleichen
  Einfaktorielle ANOVA…

Abbildung 21-1
Dialogfeld "Einfaktorielle ANOVA"

- Wählen Sie eine oder mehrere abhängige Variablen aus.
- Wählen Sie eine unabhängige Faktorvariable aus.

Einfaktorielle ANOVA: Kontraste

Abbildung 21-2
Dialogfeld "Einfaktorielle ANOVA: Kontraste"



Sie können die Quadratsummen zwischen den Gruppen in Trend-Komponenten zerlegen oder A-priori-Kontraste festlegen.

Polynomial. Damit zerlegen Sie die Quadratsummen zwischen den Gruppen in Trend-Komponenten. Sie können die abhängige Variable auf einen Trend über die geordneten Stufen der Faktorvariablen prüfen. Sie können beispielsweise prüfen, ob beim Gehalt über die

geordneten Stufen des höchsten erreichten akademischen Grads ein linearer (steigender oder fallender) Trend vorliegt.

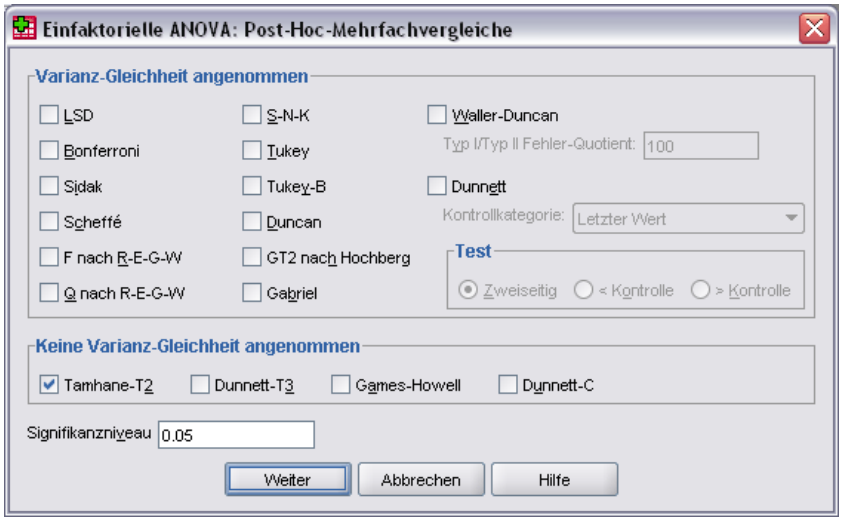
- **Grad.** Sie können Polynome ersten, zweiten, dritten, vierten und fünften Grades auswählen.

Koeffizienten. Mit der *T*-Statistik werden benutzerdefinierte A-priori-Kontraste getestet. Geben Sie für jede Gruppe (Kategorie) der Faktorvariablen einen Koeffizienten ein und klicken Sie nach jeder Eingabe auf Hinzufügen. Jeder neue Wert wird am Ende der Liste der Koeffizienten hinzugefügt. Um zusätzliche Kontrastgruppen festzulegen, klicken Sie auf Weiter. Verwenden Sie Weiter und Zurück, um zwischen den Kontrastgruppen zu wechseln.

Die Reihenfolge der Koeffizienten ist wichtig, weil sie den aufsteigend geordneten Kategoriewerten der Faktorvariablen entspricht. Der erste Koeffizient der Liste entspricht dem kleinsten Gruppenwert der Faktorvariablen, der letzte Koeffizient dem größten Wert. Bei zum Beispiel sechs Kategorien der Faktorvariablen stellen die Koeffizienten –1, 0, 0, 0,5 und 0,5 einen Kontrast zwischen der ersten und der fünften und sechsten Gruppe her. Bei den meisten Anwendungen muss die Summe der Koeffizienten 0 ergeben. Sie können auch Werte benutzen, deren Summe ungleich 0 ist. In diesem Fall wird jedoch eine Warnung angezeigt.

Einfaktorielle ANOVA: Post-Hoc-Mehrfachvergleiche

Abbildung 21-3
Dialogfeld "Einfaktorielle ANOVA: Post-Hoc-Mehrfachvergleiche"



Sobald Sie festgestellt haben, dass es Abweichungen zwischen den Mittelwerten gibt, können Sie mit Post-Hoc-Spannweiten-Tests und paarweisen multiplen Vergleichen untersuchen, welche Mittelwerte sich unterscheiden. Spannweitentests ermitteln homogene Untergruppen von Mittelwerten, die nicht voneinander abweichen. Mit paarweisen Mehrfachvergleichen testen Sie die Differenz zwischen gepaarten Mittelwerten. Die Ergebnisse werden in einer Matrix angezeigt, in der Gruppenmittelwerte, die auf einem Alpha-Niveau von 0,05 signifikant voneinander abweichen, durch Sterne markiert sind.

Varianz-Gleichheit angenommen

Die ehrlich signifikante Differenz nach Tukey, der GT2 nach Hochberg, der Gabriel-Test und der Scheffé-Test sind Tests für Mehrfachvergleiche und Spannweitentests. Andere Spannweitentests sind Tukey-*B*, S-N-K (Student-Newman-Keuls), Duncan, *F* nach R-E-G-W (*F*-Test nach Ryan-Einot-Gabriel-Welsch), *Q* nach R-E-G-W (Spannweitentest nach Ryan-Einot-Gabriel-Welsch) und Waller-Duncan. Verfügbare Tests für Mehrfachvergleiche sind Bonferroni, ehrlich signifikante Differenz nach Tukey, Sidak, Gabriel, Hochberg, Dunnett, Scheffé und LSD (geringste signifikante Differenz).

- **LSD.** Verwendet T-Tests, um alle paarweisen Vergleiche zwischen Gruppenmittelwerten durchzuführen. Es erfolgt keine Korrektur der Fehlerrate bei Mehrfachvergleichen.
- **Bonferroni.** Führt paarweise Vergleiche zwischen Gruppenmittelwerten mit T-Tests aus; regelt dabei jedoch auch die Gesamtfehlerrate, indem die Fehlerrate für jeden Test auf den Quotienten aus der experimentellen Fehlerrate und der Gesamtzahl der Tests gesetzt wird. Dadurch wird das beobachtete Signifikanzniveau für Mehrfachvergleiche angepasst.
- **Sidak.** Ein paarweiser multipler Vergleichstest, basierend auf einer T-Statistik. Beim Sidak-Test wird das Signifikanzniveau für die multiplen Vergleiche korrigiert und es werden engere Grenzen vergeben als bei Bonferroni.
- **Scheffé.** Führt gemeinsame paarweise Vergleiche gleichzeitig für alle möglichen paarweisen Kombinationen der Mittelwerte durch. Verwendet die F-Stichprobenverteilung. Dieser Test kann verwendet werden, um nicht nur paarweise Vergleiche durchzuführen, sondern alle möglichen linearen Kombinationen von Gruppenmittelwerten zu untersuchen.
- **F nach R-E-G-W.** Mehrfaches Rückschrittverfahren nach Ryan-Einot-Gabriel-Welsh, basierend auf einem F-Test.
- **Q nach R-E-G-W.** Mehrfaches Rückschrittverfahren nach Ryan-Einot-Gabriel-Welsh, basierend auf der studentisierten Spannweite.
- **S-N-K.** Führt alle paarweisen Vergleiche zwischen Mittelwerten unter Verwendung der t-Verteilung aus. Bei gleich großen Stichproben werden auch die Mittelwertpaare innerhalb homogener Untergruppen verglichen; dabei wird ein schrittweises Verfahren verwendet. Die Mittelwerte werden in absteigender Reihenfolge (vom größten zum kleinsten Wert) sortiert, extreme Differenzen werden zuerst getestet.
- **Tukey.** Verwendet die Student-Verteilung für alle möglichen paarweisen Vergleiche zwischen den Gruppen. Setzt die Fehlerrate für das Experiment gleich der Fehlerrate für die Gesamtheit aller paarweisen Vergleiche.
- **Tukey-B-Test.** Verwendet die Student-Verteilung für paarweise Vergleiche zwischen Gruppen. Der kritische Wert ist der Durchschnitt des entsprechenden Werts für die ehrlich signifikante Differenz nach Tukey und für Student-Newman-Keuls.
- **Duncan.** Bei diesem Test werden paarweise Vergleiche angestellt, deren schrittweise Reihenfolge identisch ist mit der Reihenfolge, die beim Student-Newman-Keuls-Test verwendet wird. Abweichend wird aber ein Sicherheitsniveau für die Fehlerrate der zusammengefassten Tests statt einer Fehlerrate für die einzelnen Tests festgelegt. Es wird die studentisierte Bereichsstatistik verwendet.
- **GT2 nach Hochberg.** Ein paarweiser Vergleichstest, der auf dem studentisierten Maximalmodul beruht. Ähnelt dem Test auf ehrlich signifikante Differenz nach Tukey.

- **Gabriel.** Ein paarweiser Vergleichstest, der das studentisierte Maximalmodul verwendet. Er ist in der Regel aussagekräftiger als der GT2-Test nach Hochberg, wenn unterschiedliche Zellengrößen vorliegen. Der Test nach Gabriel kann ungenau sein, wenn die Zellengrößen große Abweichungen aufweisen.
- **Waller-Duncan.** Ein Test für Mehrfachvergleiche auf der Grundlage einer T-Statistik; verwendet eine Bayes-Methode.
- **Dunnett.** Ein paarweiser T-Test für Mehrfachvergleiche, der ein Set von Verarbeitungen mit einem einzelnen Kontrollmittelwert vergleicht. Als Kontrollkategorie ist die letzte Kategorie voreingestellt. Sie können aber auch die erste Kategorie einstellen. Verwenden Sie einen zweiseitigen Test, um zu überprüfen, ob sich der Mittelwert bei jeder Stufe (außer der Kontrollkategorie) des Faktors von dem Mittelwert der Kontrollkategorie unterscheidet. Wählen Sie <>Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors kleiner als der Mittelwert der Kontrollkategorie ist. Wählen Sie >Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors größer als der Mittelwert der Kontrollkategorie ist.

Keine Varianz-Gleichheit angenommen

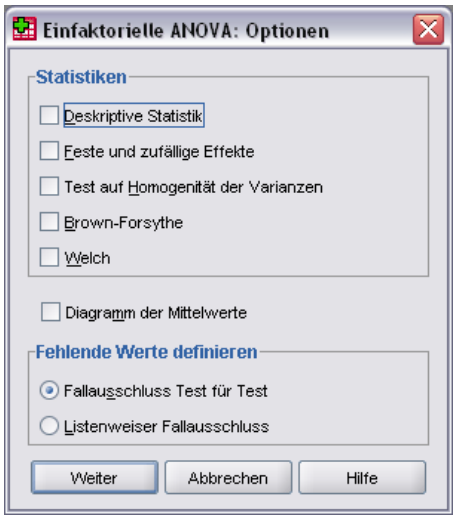
Tests für Mehrfachvergleiche, die keine Varianzgleichheit voraussetzen, sind Tamhane-T2, Dunnett-T3, Games-Howell und Dunnett-*C*.

- **T2 nach Tamhane.** Konservative, paarweise Vergleichstests auf der Grundlage eines T-Tests. Dieser Test ist für ungleiche Varianzen geeignet.
- **T3 nach Dunnett.** Ein paarweiser Vergleichstest, der auf dem studentisierten Maximalmodul beruht. Dieser Test ist für ungleiche Varianzen geeignet.
- **Games-Howell.** Ein manchmal schwacher, paarweiser Vergleichstest. Dieser Test ist für ungleiche Varianzen geeignet.
- **C nach Dunnett.** Ein paarweiser Vergleichstest, der auf dem studentisierten Bereich beruht. Dieser Test ist für ungleiche Varianzen geeignet.

Anmerkung: Die Ausgabe von Post-Hoc-Tests läßt sich oft einfacher interpretieren, wenn Sie im Dialogfeld “Tabelleneigenschaften” die Option Leere Zeilen und Spalten ausblenden deaktivieren. (In einer aktivierten Pivot-Tabelle: Tabelleneigenschaften im Menü “Format”.)

Einfaktorielle ANOVA: Optionen

Abbildung 21-4
Dialogfeld "Einfaktorielle ANOVA: Optionen"



Statistiken. Wählen Sie mindestens eine der folgenden Optionen aus:

- **Deskriptive Statistik.** Hiermit berechnen Sie Anzahl der Fälle, Mittelwert, Standardabweichung, Standardfehler des Mittelwerts, Minimum, Maximum und das 95%-Konfidenzintervall für jede abhängige Variable in jeder Gruppe.
- **Feste und zufällige Effekte.** Hiermit werden die Standardabweichung, der Standardfehler und das 95%-Konfidenzintervall für das Modell mit festen Effekten sowie der Standardfehler, das 95%-Konfidenzintervall und der Schätzer der Varianz zwischen Komponenten für das Modell mit zufälligen Effekten angezeigt.
- **Test auf Homogenität der Varianzen.** Bei dieser Option wird die Levene-Statistik berechnet, mit der Sie die Gruppenvarianzen auf Gleichheit testen können. Dieser Test setzt keine Normalverteilung voraus.
- **Brown-Forsythe.** Bei dieser Option wird die Brown-Forsythe-Statistik berechnet, mit der Sie die Gruppenmittelwerte auf Gleichheit testen können. Diese Statistik ist der *F*-Statistik vorzuziehen, wenn die Annahme gleicher Varianzen sich nicht bestätigt.
- **Welch.** Bei dieser Option wird die Welch-Statistik berechnet, mit der Sie die Gruppenmittelwerte auf Gleichheit testen können. Diese Statistik ist der *F*-Statistik vorzuziehen, wenn die Annahme gleicher Varianzen sich nicht bestätigt.

Diagramm der Mittelwerte. Bei dieser Option wird ein Diagramm für die Mittelwerte der Untergruppen ausgegeben. Dabei handelt es sich um die Mittelwerte für jede Gruppe, die durch die Werte der Faktorvariablen definiert ist.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluss Test für Test.** Bei Auswahl dieser Option werden Fälle mit einem fehlenden Wert für die abhängige Variable oder die Faktorvariable in einer bestimmten Analyse in dieser Analyse nicht verwendet. Ein Fall wird außerdem nicht verwendet, wenn er außerhalb des Bereichs liegt, der für die Faktorvariable definiert ist.
- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für die Faktorvariable oder eine abhängige Variable, die in der Liste der abhängigen Variablen des Hauptdialogfelds enthalten sind, werden aus allen Analysen ausgeschlossen. Wenn Sie nicht mehrere abhängige Variablen festgelegt haben, hat dies keine Auswirkung.

Zusätzliche Funktionen beim Befehl ONEWAY

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Erstellen von Statistiken mit festen und zufälligen Effekten. Standardabweichung, Standardfehler des Mittelwerts und 95%-Konfidenzintervalle für ein Modell mit festen Effekten. Standardfehler, 95%-Konfidenzintervalle und die Schätzung der Varianz zwischen Komponenten für ein Modell mit zufälligen Effekten (mit `STATISTICS=EFFECTS`).
- Angeben der Alpha-Niveaus für die Tests für Mehrfachvergleiche auf geringste signifikante Differenz sowie nach Bonferroni, Duncan und Scheffé (mit dem Unterbefehl `RANGES`).
- Schreiben einer Matrix der Mittelwerte, Standardabweichungen und Häufigkeiten oder Lesen einer Matrix der Mittelwerte, Häufigkeiten, gemeinsame Varianzen sowie der Freiheitsgrade für die gemeinsamen Varianzen. Diese Matrizen können anstellen der Rohdaten verwendet werden, um eine einfaktorielle Analyse der Varianz durchzuführen (mit dem Unterbefehl `MATRIX`).

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
22

GLM - Univariat

Mit der Prozedur “GLM - Univariat” können Sie Regressionsanalysen und Varianzanalysen für eine abhängige Variable mit einem oder mehreren Faktoren und/oder Variablen durchführen. Die Faktorvariablen unterteilen die Grundgesamtheit in Gruppen. Unter Verwendung dieser auf einem allgemeinen linearen Modell basierenden Prozedur können Sie Nullhypothesen über die Effekte anderer Variablen auf die Mittelwerte verschiedener Gruppierungen einer einzelnen abhängigen Variablen testen. Sie können die Wechselwirkungen zwischen Faktoren und die Effekte einzelner Faktoren untersuchen, von denen einige zufällig sein können. Außerdem können Sie die Auswirkungen von Kovariaten und Wechselwirkungen zwischen Kovariaten und Faktoren berücksichtigen. Bei der Regressionsanalyse werden die unabhängigen Variablen (Einflußvariablen) als Kovariaten angegeben.

Es können sowohl ausgeglichene als auch nicht ausgeglichene Modelle getestet werden. Ein Design ist ausgeglichen, wenn jede Zelle im Modell dieselbe Anzahl von Fällen enthält. Mit der Prozedur “GLM - Univariat” werden nicht nur Hypothesen getestet, sondern zugleich Parameter geschätzt.

Zum Testen von Hypothesen stehen häufig verwendete a-priori-Kontraste zur Verfügung. Nachdem die Signifikanz mit einem *F*-Gesamttest nachgewiesen wurde, können Sie Post-Hoc-Tests verwenden, um Differenzen zwischen bestimmten Mittelwerten berechnen zu lassen. Geschätzte Randmittel dienen als Schätzer für die vorhergesagten Mittelwerte der Zellen im Modell, und mit Profilplots (Wechselwirkungsdiagrammen) dieser Mittelwerte können Sie einige dieser Beziehungen in einfacher Weise visuell darstellen.

Residuen, Einflußwerte, die Cook-Distanz und Hebelwerte können zum Überprüfen von Annahmen als neue Variablen in der Datendatei gespeichert werden.

Mit der WLS-Gewichtung können Sie eine Variable angeben, um Beobachtungen für eine WLS-Analyse (Weighted Least Squares, deutsch: gewichtete kleinste Quadrate) unterschiedlich zu gewichten. Dies kann notwendig sein, um etwaige Unterschiede in der Präzision von Messungen auszugleichen.

Beispiel. Im Rahmen einer sportwissenschaftlichen Studie beim Berlin-Marathon werden mehrere Jahre lang Daten über einzelne Läufer aufgenommen. Die abhängige Variable ist die Zeit, die jeder Läufer für die Strecke benötigt. Andere berücksichtigte Faktoren sind beispielsweise das Wetter (kalt, angenehm oder heiß), die Anzahl von Trainingsmonaten, die Anzahl der bereits absolvierten Marathons und das Geschlecht. Das Alter der betreffenden Personen wird als Kovariate betrachtet. Ein mögliches Ergebnis wäre, dass das Geschlecht ein signifikanter Effekt und die Wechselwirkung von Geschlecht und Wetter signifikant ist.

Methoden. Zum Überprüfen der verschiedenen Hypothesen können Quadratsummen vom Typ I, Typ II, Typ III und Typ IV verwendet werden. Die Voreinstellung sieht den Typ III vor.

Statistik. Post-Hoc-Spannweitentests und Mehrfachvergleiche: geringste signifikante Differenz, Bonferroni, Sidak, Scheffé, multiples *F* nach Ryan-Einot-Gabriel-Welsch, multiple Spannweite nach Ryan-Einot-Gabriel-Welsch, Student-Newman-Keuls-Test, ehrlich signifikante Differenz nach Tukey, Tukey-*B*, Duncan, GT2 nach Hochberg, Gabriel, Waller-Duncan-*T*-Test, Dunnett (einseitig und zweiseitig), Tamhane-T2, Dunnett-T3, Games-Howell und Dunnett-*C*. Deskriptive Statistiken: beobachtete Mittelwerte, Standardabweichungen und Häufigkeiten aller abhängigen Variablen in allen Zellen. Levene-Test auf Homogenität der Varianzen.

Diagramme. Streubreite gegen mittleres Niveau, Residuen-Diagramme, Profilplots (Wechselwirkung).

Daten. Die abhängige Variable ist quantitativ. Faktoren sind kategorial. Sie können numerische Werte oder String-Werte von bis zu acht Zeichen Länge annehmen. Kovariaten sind quantitative Variablen, die mit der abhängigen Variablen in Beziehung stehen.

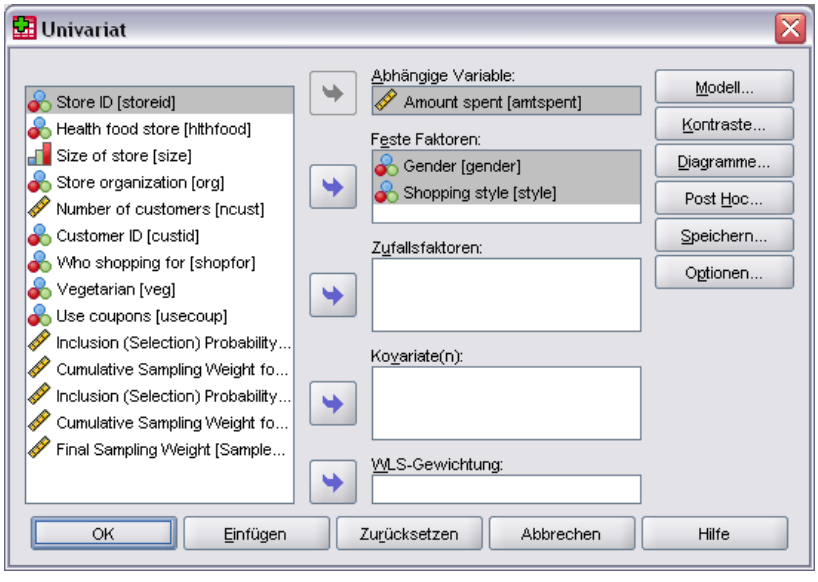
Annahmen. Die Daten sind eine Stichprobe aus einer normalverteilten Grundgesamtheit. In der Grundgesamtheit sind alle Zellenvarianzen gleich. Die Varianzanalyse ist unempfindlich gegenüber Abweichungen von der Normalverteilung. Die Daten müssen jedoch symmetrisch verteilt sein. Zum Überprüfen der Annahmen können Sie Tests auf Homogenität der Varianzen vornehmen und Diagramme der Streubreite gegen das mittlere Niveau ausgeben lassen. Sie können auch die Residuen untersuchen und Residuen-Diagramme anzeigen lassen.

So berechnen Sie eine univariate Analyse der Varianz (GLM):

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Allgemeines lineares Modell
    Univariat...

Abbildung 22-1
Dialogfeld "GLM - Univariat"

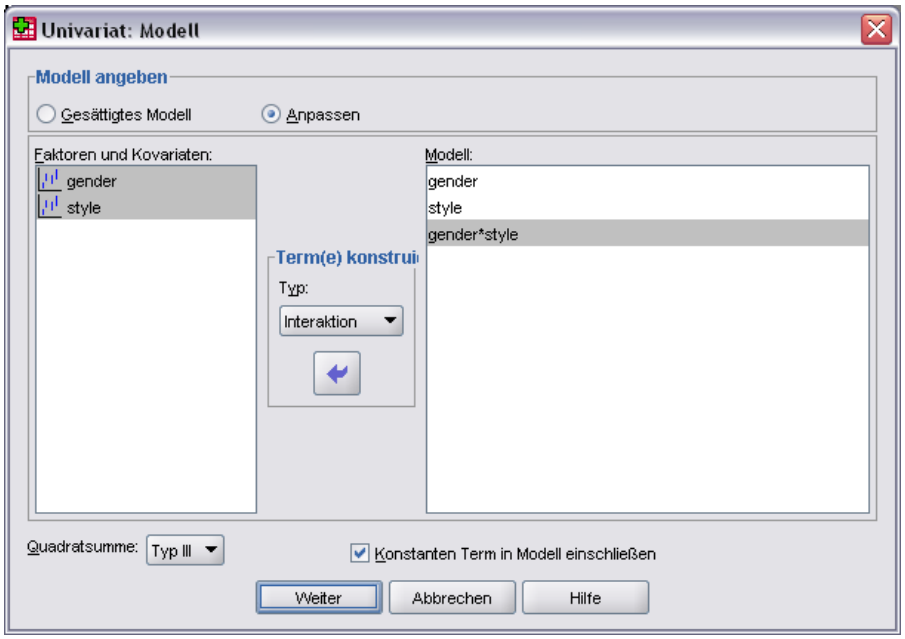


▶ Wählen Sie eine abhängige Variable aus.

- Wählen Sie in Abhängigkeit von den Daten Variablen als feste Faktoren, Zufallsfaktoren und Kovariaten aus.
- Optional können Sie mit der WLS-Gewichtung eine Gewichtungsvariable für WLS-Analyse (Weighted Least Squares, gewichtete kleinste Quadrate) angeben. Wenn der Wert der Gewichtungsvariablen null, negativ oder fehlend ist, wird der Fall aus der Analyse ausgeschlossen. Eine bereits im Model verwendete Variable kann nicht als Gewichtungsvariable verwendet werden.

GLM: Modell

Abbildung 22-2
Dialogfeld "Univariat: Modell"



Modell angeben. Ein gesättigtes Modell enthält alle Faktoren-Haupteffekte, alle Kovariaten-Haupteffekte und alle faktorweisen Wechselwirkungen. Es enthält keine Kovariaten-Wechselwirkungen. Wählen Sie Anpassen aus, um nur eine Teilmenge von Wechselwirkungen oder Wechselwirkungen zwischen Faktoren und Kovariaten festzulegen. Sie müssen alle in das Modell zu übernehmenden Terme angeben.

Faktoren und Kovariaten. Die Faktoren und Kovariaten werden aufgelistet.

Modell. Das Modell ist von der Art Ihrer Daten abhängig. Nach der Auswahl von Anpassen können Sie die Haupteffekte und Wechselwirkungen auswählen, die für Ihre Analyse von Interesse sind.

Quadratsumme. Hier wird die Methode zum Berechnen der Quadratsumme festgelegt. Für ausgeglichene und unausgeglichene Modelle ohne fehlende Zellen wird meistens die Methode mit Quadratsummen vom Typ III angewendet.

Konstanten Term in Modell einschließen. Der konstante Term wird gewöhnlich in das Modell aufgenommen. Falls Sie sicher sind, dass die Daten durch den Koordinatenursprung verlaufen, können Sie den konstanten Term ausschließen.

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-fach. Hiermit werden alle möglichen 2-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-fach. Hiermit werden alle möglichen 3-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-fach. Hiermit werden alle möglichen 4-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-fach. Hiermit werden alle möglichen 5-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Quadratsumme

Für das Modell können Sie einen Typ von Quadratsumme auswählen. Typ III wird am häufigsten verwendet und ist die Standardeinstellung.

Typ I. Diese Methode ist auch als die Methode der hierarchischen Zerlegung der Quadratsummen bekannt. Jeder Term wird nur für den Vorläuferterm im Modell angepaßt. Quadratsummen vom Typ I werden gewöhnlich in den folgenden Situationen verwendet:

- Ein ausgeglichenes ANOVA-Modell, in dem alle Haupteffekte vor den Wechselwirkungseffekten 1. Ordnung festgelegt werden, alle Wechselwirkungseffekte 1. Ordnung wiederum vor den Wechselwirkungseffekten 2. Ordnung festgelegt werden und so weiter.
- Ein polynomiales Regressionsmodell, in dem alle Terme niedrigerer Ordnung vor den Termen höherer Ordnung festgelegt werden.
- Ein rein verschachteltes Modell, in welchem der zuerst bestimmte Effekt in dem als zweiten bestimmten Effekt verschachtelt ist, der zweite Effekt wiederum im dritten und so weiter. (Diese Form der Verschachtelung kann nur durch Verwendung der Befehlssprache erreicht werden.)

Typ II. Bei dieser Methode wird die Quadratsumme eines Effekts im Modell angepaßt an alle anderen "zutreffenden" Effekte berechnet. Ein zutreffender Effekt ist ein Effekt, der mit allen Effekten in Beziehung steht, die den untersuchten Effekt nicht enthalten. Die Methode mit Quadratsummen vom Typ II wird gewöhnlich in den folgenden Fällen verwendet:

- Bei ausgeglichenen ANOVA-Modellen.
- Bei Modellen, die nur Haupteffekte von Faktoren enthalten.
- Bei Regressionsmodellen.
- Bei rein verschachtelten Designs. (Diese Form der Verschachtelung kann durch Verwendung der Befehlssprache erreicht werden.)

Typ III. Voreinstellung. Bei dieser Methode werden die Quadratsummen eines Effekts im Design als Quadratsummen orthogonal zu den Effekten (sofern vorhanden), die den Effekt enthalten, und mit Bereinigung um alle anderen Effekte, die diesen Effekt nicht enthalten, berechnet. Der große Vorteil der Quadratsummen vom Typ III ist, daß sie invariant bezüglich der Zellenhäufigkeiten sind, solange die allgemeine Form der Schätzbarkeit konstant bleibt. Daher wird dieser Typ von Quadratsumme oft für nicht ausgeglichene Modelle ohne fehlende Zellen als geeignet angesehen. In einem faktoriellen Design ohne fehlende Zellen ist diese Methode äquivalent zu der Methode der gewichteten Mittelwertquadrate nach Yates. Die Methode mit Quadratsummen vom Typ III wird gewöhnlich in folgenden Fällen verwendet:

- Alle bei Typ I und Typ II aufgeführten Modelle.
- Alle ausgeglichenen oder unausgeglichenen Modelle ohne leere Zellen.

Typ IV. Diese Methode ist dann geeignet, wenn es keine fehlenden Zellen gibt. Für alle Effekte *F* im Design: Wenn *F* in keinem anderen Effekt enthalten ist, dann gilt: Typ IV = Typ III = Typ II. Wenn *F* in anderen Effekten enthalten ist, werden bei Typ IV die Kontraste zwischen den Parametern in *F* gleichmäßig auf alle Effekte höherer Ordnung verteilt. Die Methode mit Quadratsummen vom Typ IV wird gewöhnlich in folgenden Fällen verwendet:

- Alle bei Typ I und Typ II aufgeführten Modelle.
- Alle ausgeglichenen oder unausgeglichenen Modelle mit leeren Zellen.

GLM: Kontraste

Abbildung 22-3
Dialogfeld "Univariat: Kontraste"

Kontraste werden verwendet, um auf Unterschiede zwischen den Stufen eines Faktors zu testen. Für jeden Faktor im Modell kann ein Kontrast festgelegt werden (in einem Modell mit Messwiederholungen für jeden Zwischensubjektfaktor). Kontraste stellen lineare Kombinationen der Parameter dar.

Das Testen der Hypothesen basiert auf der Nullhypothese **LB** = 0. Dabei ist **L** die Kontrastkoeffizienten-Matrix und **B** der Parametervektor. Wenn ein Kontrast angegeben wird, wird eine **L**-Matrix erstellt. Die Spalten der **L**-Matrix, die dem Faktor entsprechen, stimmen mit dem Kontrast überein. Die verbleibenden Spalten werden so angepaßt, dass die **L**-Matrix schätzbar ist.

Die Ausgabe beinhaltet eine *F*-Statistik für jedes Set von Kontrasten. Für die Kontrastdifferenzen werden außerdem simultane Konfidenzintervalle nach Bonferroni auf der Grundlage der Student-*T*-Verteilung angezeigt.

Verfügbare Kontraste

Als Kontraste sind “Abweichung”, “Einfach”, “Differenz”, “Helmert”, “Wiederholt” und “Polynomial” verfügbar. Bei Abweichungskontrasten und einfachen Kontrasten können Sie wählen, ob die letzte oder die erste Kategorie als Referenzkategorie dient.

Kontrasttypen

Abweichung. Vergleicht den Mittelwert jeder Faktorstufe (außer bei Referenzkategorien) mit dem Mittelwert aller Faktorstufen (Gesamtmittelwert). Die Stufen des Faktors können in beliebiger Ordnung vorliegen.

Einfach. Vergleicht den Mittelwert jeder Faktorstufe mit dem Mittelwert einer angegebenen Faktorstufe. Dieser Kontrasttyp ist nützlich, wenn es eine Kontrollgruppe gibt. Sie können die erste oder die letzte Kategorie als Referenz auswählen.

Differenz. Vergleicht den Mittelwert jeder Faktorstufe (außer der ersten) mit dem Mittelwert der vorhergehenden Faktorstufen. (Dies wird gelegentlich auch als umgekehrter Helmert-Kontrast bezeichnet).

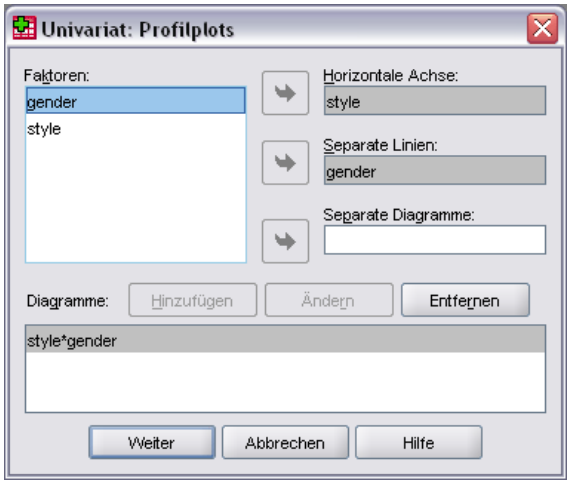
Helmert. Vergleicht den Mittelwert jeder Stufe des Faktors (bis auf die letzte) mit dem Mittelwert der folgenden Stufen.

Wiederholt. Vergleicht den Mittelwert jeder Faktorstufe (außer der letzten) mit dem Mittelwert der folgenden Faktorstufe.

Polynomial. Vergleicht den linearen Effekt, quadratischen Effekt, kubischen Effekt und so weiter. Der erste Freiheitsgrad enthält den linearen Effekt über alle Kategorien; der zweite Freiheitsgrad den quadratischen Effekt und so weiter. Die Kontraste werden oft verwendet, um polynomiale Trends zu schätzen.

GLM: Profilplots

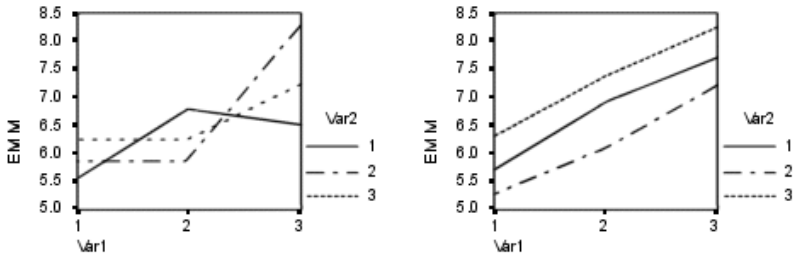
Abbildung 22-4
Dialogfeld "Univariat: Profilplots"



Profilplots (Wechselwirkungsdiagramme) sind hilfreich zum Vergleichen von Randmitteln im Modell. Ein Profilplot ist ein Liniendiagramm, in dem jeder Punkt das geschätzte Randmittel einer abhängigen Variablen (angepaßt an die Kovariaten) bei einer Stufe eines Faktors angibt. Die Stufen eines zweiten Faktors können zum Erzeugen getrennter Linien verwendet werden. Jede Stufe in einem dritten Faktor kann verwendet werden, um ein separates Diagramm zu erzeugen. Alle festen Faktoren und Zufallsfaktoren (sofern vorhanden) sind für Diagramme verfügbar. Bei multivariaten Analysen werden Profilplots für jede abhängige Variable erstellt. Bei einer Analyse mit Messwiederholungen können in Profilplots sowohl Zwischensubjektfaktoren als auch Innersubjektfaktoren verwendet werden. "GLM – Multivariat" und "GLM – Messwiederholungen" sind nur verfügbar, wenn Sie SPSS Advanced Models installiert haben.

Ein Profilplot für einen Faktor zeigt, ob die geschätzten Randmittel mit den Faktorstufen steigen oder fallen. Bei zwei oder mehr Faktoren deuten parallele Linien an, dass es keine Wechselwirkung zwischen den Faktoren gibt. Das heißt, dass Sie die Faktorstufen eines einzelnen Faktors untersuchen können. Nichtparallele Linien deuten auf eine Wechselwirkung hin.

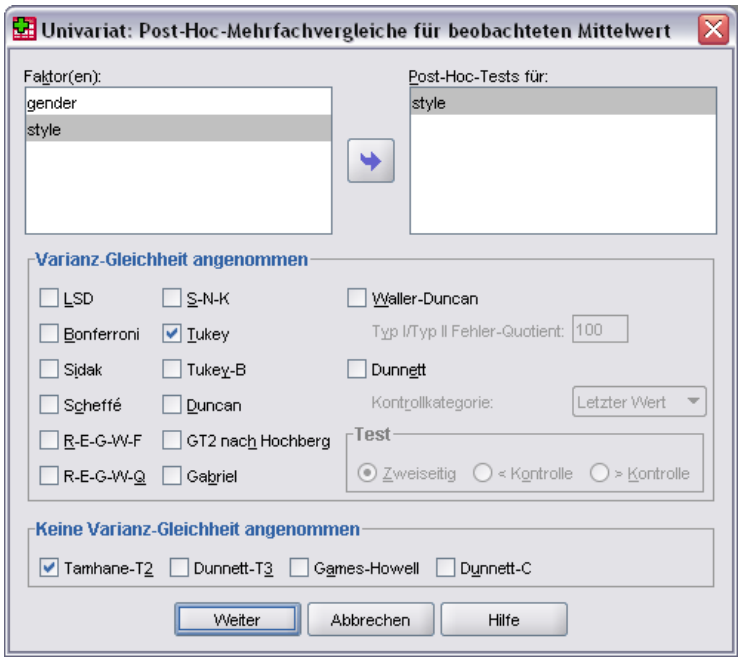
Abbildung 22-5
Nichtparalleles Diagramm (links) und paralleles Diagramm (rechts)



Nachdem ein Diagramm durch Auswahl von Faktoren für die horizontale Achse (und wahlweise von Faktoren für getrennte Linien und getrennte Diagramme) festgelegt wurde, muss das Diagramm der Liste "Diagramme" hinzugefügt werden.

GLM: Post-Hoc-Vergleiche

Abbildung 22-6
Dialogfeld "Post Hoc"



Tests für Post-Hoc-Mehrfachvergleiche. Sobald Sie festgestellt haben, daß es Abweichungen zwischen den Mittelwerten gibt, können Sie mit Post-Hoc-Spannweiten-Tests und paarweisen multiplen Vergleichen untersuchen, welche Mittelwerte sich unterscheiden. Die Vergleiche werden auf der Basis von nicht korrigierten Werten vorgenommen. Diese Tests werden nur für feste Zwischensubjektfaktoren durchgeführt. Bei "GLM - Meßwiederholungen" sind diese Tests nicht verfügbar, wenn es keine Zwischensubjektfaktoren gibt, und die Post-Hoc-Mehrfachvergleiche werden für den Durchschnitt aller Stufen der Innersubjektfaktoren durchgeführt. Bei "GLM - Multivariat" werden für jede abhängige Variable eigene Post-Hoc-Tests durchgeführt. "GLM – Multivariat" und "GLM – Meßwiederholungen" sind nur verfügbar, wenn Sie SPSS Advanced Models installiert haben.

Häufig verwendete Mehrfachvergleiche sind der Bonferroni-Test und die ehrlich signifikante Differenz nach Tukey. Der **Bonferroni-Test** auf der Grundlage der studentisierten *T*-Statistik korrigiert das beobachtete Signifikanzniveau unter Berücksichtigung der Tatsache, daß multiple Vergleiche vorgenommen werden. Der **Sidak-T-Test** korrigiert ebenfalls das Signifikanzniveau und liefert engere Grenzen als der Bonferroni-Test. Die **ehrlich signifikante Differenz nach Tukey** verwendet die studentisierte Spannweitenstatistik, um alle paarweisen Vergleiche zwischen den Gruppen vorzunehmen, und setzt die experimentelle Fehlerrate auf die Fehlerrate der Ermittlung aller paarweisen Vergleiche. Beim Testen einer großen Anzahl von Mittelwertpaaren ist der Test auf ehrlich signifikante Differenz nach Tukey leistungsfähiger als der Bonferroni-Test. Bei einer kleinen Anzahl von Paaren ist der Bonferroni-Test leistungsfähiger.

GT2 nach Hochberg ähnelt dem Test auf ehrlich signifikante Differenz nach Tukey, es wird jedoch das studentisierte Maximalmodul verwendet. Meistens ist der Test nach Tukey leistungsfähiger. Der **paarweise Vergleichstest nach Gabriel** verwendet ebenfalls

das studentisierte Maximalmodul und zeigt meistens eine größere Schärfe als das GT2 nach Hochberg, wenn die Zellengrößen ungleich sind. Der Gabriel-Test kann ungenau werden, wenn die Zellengrößen stark variieren.

Mit dem **paarweisen T-Test für mehrere Vergleiche nach Dunnett** wird ein Set von Verarbeitungen mit einem einzelnen Kontrollmittelwert verglichen. Als Kontrollkategorie ist die letzte Kategorie voreingestellt. Sie können aber auch die erste Kategorie einstellen. Außerdem können Sie einen einseitigen oder zweiseitigen Test wählen. Verwenden Sie einen zweiseitigen Test, um zu überprüfen, ob sich der Mittelwert bei jeder Stufe (außer der Kontrollkategorie) des Faktors von dem Mittelwert der Kontrollkategorie unterscheidet. Wählen Sie < Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors kleiner als der Mittelwert der Kontrollkategorie ist. Wählen Sie > Kontrolle, um zu überprüfen, ob der Mittelwert bei allen Stufen des Faktors größer als der Mittelwert bei der Kontrollkategorie ist.

Ryan, Einot, Gabriel und Welsch (R-E-G-W) entwickelten zwei multiple Step-Down-Spannweitentests. Multiple Step-Down-Prozeduren überprüfen zuerst, ob alle Mittelwerte gleich sind. Wenn nicht alle Mittelwerte gleich sind, werden Teilmengen der Mittelwerte auf Gleichheit getestet. Das **F nach R-E-G-W** basiert auf einem *F*-Test, und **Q nach R-E-G-W** basiert auf der studentisierten Spannweite. Diese Tests sind leistungsfähiger als der multiple Spannweitentest nach Duncan und der Student-Newman-Keuls-Test (ebenfalls multiple Step-Down-Prozeduren), aber sie sind bei ungleichen Zellengrößen nicht empfehlenswert.

Bei ungleichen Varianzen verwenden Sie das **Tamhane-T2** (konservativer paarweiser Vergleichstest auf der Grundlage eines *T*-Tests), **Dunnett-T3** (paarweiser Vergleichstest auf der Grundlage des studentisierten Maximalmoduls), den **paarweisen Vergleichstest** nach **Games-Howell** (manchmal ungenau) oder das **Dunnett-C** (paarweiser Vergleichstest auf der Grundlage der studentisierten Spannweite).

Der **multiple Spannweitentest nach Duncan**, Student-Newman-Keuls (**S-N-K**) und **Tukey-B** sind Spannweitentests, mit denen Mittelwerte von Gruppen geordnet und ein Wertebereich berechnet wird. Diese Tests werden nicht so häufig verwendet wie die vorher beschriebenen Tests.

Der **Waller-Duncan-T-Test** verwendet die Bayes-Methode. Dieser Spannweitentest verwendet den harmonischen Mittelwert der Stichprobengröße, wenn die Stichprobengrößen ungleich sind.

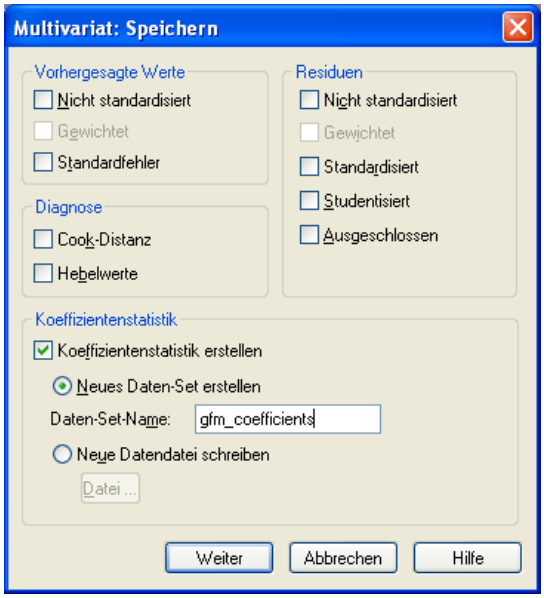
Das Signifikanzniveau des **Scheffé**-Tests ist so festgelegt, dass alle möglichen linearen Kombinationen von Gruppenmittelwerten getestet werden können und nicht nur paarweise Vergleiche verfügbar sind, wie bei dieser Funktion der Fall. Das führt dazu, dass der Scheffé-Test oftmals konservativer als andere Tests ist, also für eine Signifikanz eine größere Differenz der Mittelwerte erforderlich ist.

Der paarweise multiple Vergleichstest auf geringste signifikante Differenz (**LSD**) ist äquivalent zu multiplen individuellen *T*-Tests zwischen allen Gruppenpaaren. Der Nachteil bei diesem Test ist, daß kein Versuch unternommen wird, das beobachtete Signifikanzniveau im Hinblick auf multiple Vergleiche zu korrigieren.

Angezeigte Tests. Es werden paarweise Vergleiche für LSD, Sidak, Bonferroni, Games-Howell, T2 und T3 nach Tamhane, Dunnett-*C* und Dunnett-T3 ausgegeben. Homogene Untergruppen für Spannweitentests werden ausgegeben für S-N-K, Tukey-*B*, Duncan, *F* nach R-E-G-W, *Q* nach R-E-G-W und Waller. Die ehrlich signifikante Differenz nach Tukey, das GT2 nach Hochberg, der Gabriel-Test und der Scheffé-Test sind multiple Vergleiche, zugleich aber auch Spannweitentests.

GLM: Speichern

Abbildung 22-7
Dialogfeld "Speichern"



Vom Modell vorhergesagte Werte, Residuen und verwandte Maße können als neue Variablen im Daten-Editor gespeichert werden. Viele dieser Variablen können zum Untersuchen von Annahmen über die Daten verwendet werden. Um die Werte zur Verwendung in einer anderen SPSS-Sitzung zu speichern, müssen Sie die aktuelle Datendatei speichern.

Vorhergesagte Werte. Dies sind die Werte, welche das Modell für jeden Fall vorhersagt.

- **Nicht standardisiert.** Der Wert, den das Modell für die abhängige Variable vorhersagt.
- **Gewichtet.** Gewichtete nichtstandardisierte vorhergesagte Werte. Nur verfügbar, wenn zuvor eine WLS-Variable ausgewählt wurde.
- **Standardfehler.** Ein Schätzer der Standardabweichung des Durchschnittswerts der abhängigen Variablen für die Fälle, die dieselben Werte für die unabhängigen Variablen haben.

Diagnose. Dies sind Maße zum Auffinden von Fällen mit ungewöhnlichen Wertekombinationen bei der unabhängigen Variablen und von Fällen, die einen großen Einfluß auf das Modell haben könnten.

- **Cook-Distanz.** Ein Maß dafür, wie stark sich die Residuen aller Fälle ändern würden, wenn ein spezieller Fall von der Berechnung der Regressionskoeffizienten ausgeschlossen würde. Ein großer Wert der Cook-Distanz zeigt an, dass der Ausschluss eines Falles von der Berechnung der Regressionskoeffizienten die Koeffizienten substanziell verändert.
- **Hebelwerte.** Nicht zentrierte Hebelwerte. Der relative Einfluss einer jeden Beobachtung auf die Anpassungsgüte eines Modells.

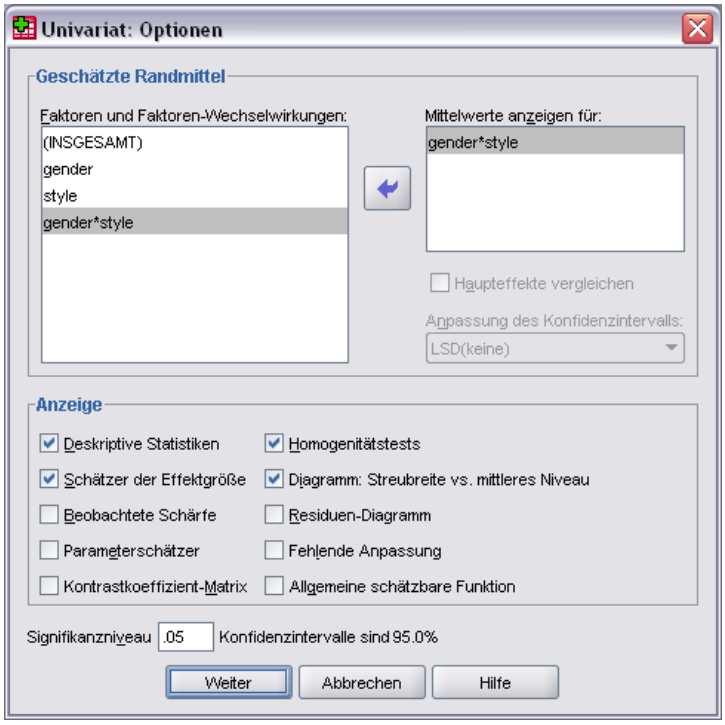
Residuen. Ein nicht standardisiertes Residuum ist der tatsächliche Wert der abhängigen Variablen minus des vom Modell geschätzten Werts. Ebenfalls verfügbar sind standardisierte, studentisierte und ausgeschlossene Residuen. Falls Sie eine WLS-Variable ausgewählt haben, sind auch gewichtete nicht standardisierte Residuen verfügbar.

- **Nicht standardisiert.** Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.
- **Gewichtet.** Gewichtete nichtstandardisierte Residuen. Nur verfügbar, wenn zuvor eine WLS-Variable ausgewählt wurde.
- **Standardisiert.** Der Quotient aus dem Residuum und einem Schätzer seiner Standardabweichung. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- **Studentisiert.** Ein Residuum, das durch seine geschätzte Standardabweichung geteilt wird, die je nach der Distanz zwischen den Werten der unabhängigen Variablen des Falles und dem Mittelwert der unabhängigen Variablen von Fall zu Fall variiert.
- **Ausgeschlossen.** Das Residuum für einen Fall, wenn dieser Fall nicht in die Berechnung der Regressionskoeffizienten eingegangen ist. Es ist die Differenz zwischen dem Wert der abhängigen Variablen und dem korrigierten Schätzwert.

Koeffizientenstatistik. Hiermit wird eine Varianz-Kovarianz-Matrix der Parameterschätzungen für das Modell in ein neues Daten-Set in der aktuellen Sitzung oder in eine externe Datei im SPSS-Format geschrieben. Für jede abhängige Variable gibt es weiterhin eine Zeile mit Parameterschätzungen, eine Zeile mit Signifikanzwerten für die *T*-Statistik der betreffenden Parameterschätzungen und eine Zeile mit den Freiheitsgraden der Residuen. Bei multivariaten Modellen gibt es ähnliche Zeilen für jede abhängige Variable. Sie können diese Matrixdatei auch in anderen Prozeduren verwenden, die Matrixdateien einlesen.

GLM: Optionen

Abbildung 22-8
Dialogfeld "Optionen"



In diesem Dialogfeld sind weitere Statistiken verfügbar. Diese werden auf der Grundlage eines Modells mit festen Effekten berechnet.

Geschätzte Randmittel. Wählen Sie die Faktoren und Wechselwirkungen aus, für die Sie Schätzer für die Randmittel der Grundgesamtheit in den Zellen wünschen. Diese Mittel werden gegebenenfalls an die Kovariaten angepaßt.

- **Haupteffekte vergleichen.** Gibt nicht korrigierte paarweise Vergleiche zwischen den geschätzten Randmitteln für alle Haupteffekte im Modell aus, sowohl für Zwischensubjektfaktoren als auch für Innersubjektfaktoren. Diese Option ist nur verfügbar, falls in der Liste "Mittelwerte anzeigen für" Haupteffekte ausgewählt sind.
- **Anpassung des Konfidenzintervalls.** Wählen Sie für das Konfidenzintervall und die Signifikanz entweder die geringste signifikante Differenz (LSD; least significant difference), Bonferroni oder die Anpassung nach Sidak. Diese Option ist nur verfügbar, wenn Haupteffekte vergleichen ausgewählt ist.

Anzeigen. Mit der Option Deskriptive Statistik lassen Sie beobachtete Mittelwerte, Standardabweichungen und Häufigkeiten für alle abhängigen Variablen in allen Zellen berechnen. Die Option Schätzer der Effektgröße liefert einen partiellen Eta-Quadrat-Wert für jeden Effekt und jede Parameterschätzung. Die Eta-Quadrat-Statistik beschreibt den Anteil der Gesamtvariabilität, der einem Faktor zugeschrieben werden kann. Die Option Beobachtete Schärfe liefert die Testschärfe, wenn die alternative Hypothese auf die Basis der beobachteten Werte eingestellt wurde. Mit Parameterschätzer werden Parameterschätzer, Standardfehler, *T*-Tests, Konfidenzintervalle und die beobachtete Schärfe für jeden Test berechnet. Mit der Option Matrix-Kontrastkoeffizienten wird die **L**-Matrix berechnet.

Mit der Option Homogenitätstest wird der Levene-Test auf Homogenität der Varianzen für alle abhängigen Variablen über alle Kombinationen von Faktorstufen der Zwischensubjektfaktoren durchgeführt (nur für Zwischensubjektfaktoren). Die Optionen für Diagramme der Streubreite gegen das mittlere Niveau und Residuen-Diagramme sind beim Überprüfen von Annahmen über die Daten nützlich. Diese Option ist nur verfügbar, wenn Faktoren vorhanden sind. Wählen Sie Residuen-Diagramm, wenn Sie für jede abhängige Variable ein Residuen-Diagramm (beobachtete über vorhergesagte über standardisierte Werte) erhalten möchten. Diese Diagramme sind beim Überprüfen der Annahme von Gleichheit der Varianzen nützlich. Mit der Option Fehlende Anpassung können Sie überprüfen, ob das Modell die Beziehung zwischen der abhängigen Variablen und der unabhängigen Variablen richtig beschreiben kann. Die Option Allgemeine schätzbare Funktion ermöglicht Ihnen, einen benutzerdefinierten Hypothesentest zu entwickeln, dessen Grundlage die allgemeine schätzbare Funktion ist. Zeilen in einer beliebigen Matrix der Kontrastkoeffizienten sind lineare Kombinationen der allgemeinen schätzbaren Funktion.

Signifikanzniveau. Hier können Sie das in den Post-Hoc-Tests verwendete Signifikanzniveau und das beim Berechnen von Konfidenzintervallen verwendete Konfidenzniveau ändern. Der hier festgelegte Wert wird auch zum Berechnen der beobachteten Schärfe für die Tests verwendet. Wenn Sie ein Signifikanzniveau festlegen, wird das entsprechende Konfidenzniveau im Dialogfeld angezeigt.

Zusätzliche Funktionen beim Befehl UNIANOVA

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Mit dem Unterbefehl `DESIGN` können Sie verschachtelte Effekte im Design festlegen.
- Mit dem Unterbefehl `TEST` können Sie Tests auf Effekte im Vergleich zu linearen Kombinationen von Effekten oder einem Wert vornehmen.
- Mit dem Unterbefehl `CONTRAST` können Sie multiple Kontraste angeben.
- Mit dem Unterbefehl `MISSING` können Sie benutzerdefinierte fehlende Werte aufnehmen.
- Mit dem Unterbefehl `CRITERIA` können Sie EPS-Kriterien angeben.
- Mit den Unterbefehlen `LMATRIX`, `MMATRIX` und `KMATRIX` können Sie benutzerdefinierte **L**-Matrizen, **M**-Matrizen und **K**-Matrizen erstellen.
- Mit dem Unterbefehl `CONTRAST` können Sie bei einfachen und Abweichungskontrasten eine Referenzkategorie zwischenschalten.
- Mit dem Unterbefehl `CONTRAST` können Sie bei polynomialen Kontrasten Metriken angeben.
- Mit dem Unterbefehl `POSTHOC` können Sie Fehlerterme für Post-Hoc-Vergleiche angeben.
- Mit dem Unterbefehl `EMMEANS` können Sie geschätzte Randmittel für alle Faktoren oder Faktorenwechselwirkungen zwischen den Faktoren in der Faktorenliste berechnen lassen.
- Mit dem Unterbefehl `SAVE` können Sie Namen für temporäre Variablen angeben.
- Mit dem Unterbefehl `OUTFILE` können Sie eine Datendatei mit einer Korrelationsmatrix erstellen.
- Mit dem Unterbefehl `OUTFILE` können Sie eine Matrix-Datendatei erstellen, die Statistiken aus der Zwischensubjekt-ANOVA-Tabelle enthält.
- Mit dem Unterbefehl `OUTFILE` können Sie die Design-Matrix in einer neuen Datendatei speichern.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
23

Bivariate Korrelationen

Mit der Prozedur "Bivariate Korrelationen" werden der Korrelationskoeffizient nach Pearson, Spearman-Rho und Kendall-Tau-*b* mit ihren jeweiligen Signifikanzniveaus errechnet. Mit Korrelationen werden die Beziehungen zwischen Variablen oder deren Rängen gemessen. Untersuchen Sie Ihre Daten vor dem Berechnen eines Korrelationskoeffizienten auf Ausreißer, da diese zu irreführenden Ergebnissen führen können. Stellen Sie fest, ob wirklich ein linearer Zusammenhang existiert. Der Korrelationskoeffizient nach Pearson ist ein Maß für den linearen Zusammenhang. Wenn zwei Variablen miteinander in starker Beziehung stehen, der Zusammenhang aber nicht linear ist, ist der Korrelationskoeffizient nach Pearson keine geeignete Statistik zum Messen des Zusammenhangs.

Beispiel. Besteht eine Korrelation zwischen der Anzahl der von einer Basketballmannschaft gewonnenen Spiele und der durchschnittlich pro Spiel erzielten Anzahl von Punkten? Ein Streudiagramm zeigt, dass ein linearer Zusammenhang besteht. Eine Analyse der Daten der NBA-Saison 1994–1995 ergibt, dass der Korrelationskoeffizient nach Pearson (0,581) auf dem Niveau 0,01 signifikant ist. Man könnte vermuten, dass die gegnerischen Mannschaften um so weniger Punkte erreicht haben, je mehr Spiele eine Mannschaft gewann. Zwischen diesen Variablen besteht eine negative Korrelation (–0,401), die auf dem Niveau 0,05 signifikant ist.

Statistiken. Für jede Variable: Anzahl der Fälle mit nichtfehlenden Werten, Mittelwert und Standardabweichung. Für jedes Variablenpaar: Korrelationskoeffizient nach Pearson, Spearman-Rho, Kendall-Tau-*b*, Kreuzprodukt der Abweichungen und Kovarianz.

Daten. Verwenden Sie symmetrische quantitative Variablen für den Korrelationskoeffizienten nach Pearson und quantitative Variablen oder Variablen mit ordinalskalierten Kategorien für das Spearman-Rho und Kendall-Tau-*b*.

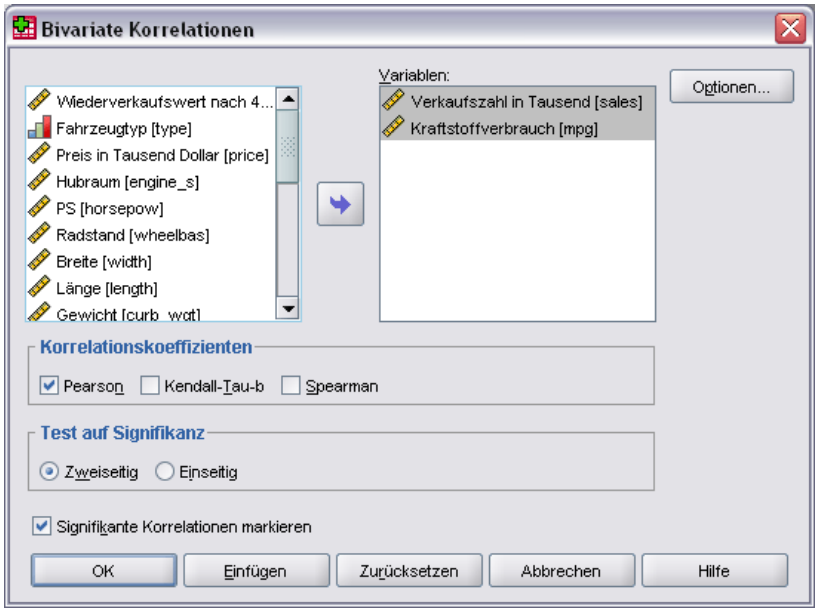
Annahmen. Für den Korrelationskoeffizient nach Pearson wird angenommen, dass jedes Variablenpaar bivariat normalverteilt ist.

So lassen Sie bivariate Korrelationen berechnen:

Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Korrelation
    Bivariat...

Abbildung 23-1
Dialogfeld "Bivariate Korrelationen"



▶ Wählen Sie mindestens zwei numerische Variablen aus.

Außerdem sind folgende Optionen verfügbar:

- **Korrelationskoeffizienten.** Für quantitative, normalverteilte Variablen wählen Sie den Korrelationskoeffizienten nach Pearson. Wenn ihre Daten nicht normalverteilt sind oder mit geordneten Kategorien vorliegen, wählen Sie die Methoden Kendall-Tau-b oder Spearman, mit denen die Beziehungen zwischen Rangordnungen gemessen werden. Der Wertebereich für Korrelationskoeffizienten reicht von –1 (perfekter negativer Zusammenhang) bis +1 (perfekter positiver Zusammenhang). Der Wert 0 bedeutet, dass kein linearer Zusammenhang besteht. Vermeiden Sie bei der Interpretation Ihrer Ergebnisse, Schlüsse über Ursache und Wirkung aufgrund signifikanter Korrelationen zu ziehen.
- **Test auf Signifikanz.** Sie können einseitige oder zweiseitige Wahrscheinlichkeiten wählen. Wenn Ihnen die Richtung des Zusammenhangs im voraus bekannt ist, wählen Sie Einseitig. Wählen Sie anderenfalls Zweiseitig.
- **Signifikante Korrelationen markieren.** Korrelationskoeffizienten, die signifikant auf dem 0,05-Niveau liegen, werden mit einem einfachen Stern angezeigt. Liegen diese signifikant auf dem 0,01-Niveau, werden sie mit zwei Sternen angezeigt.

Bivariate Korrelationen: Optionen

Abbildung 23-2
Dialogfeld "Bivariate Korrelationen: Optionen"

Statistik. Für Pearson-Korrelationen können Sie eine oder auch beide der folgenden Optionen wählen:

- **Mittelwerte und Standardabweichungen.** Diese werden für jede Variable angezeigt. Außerdem wird die Anzahl der Fälle mit nichtfehlenden Werten angezeigt. Fehlende Werte werden Variable für Variable bearbeitet, unabhängig von Ihren Einstellungen für fehlende Werte.
- **Kreuzproduktabweichungen und Kovarianzen.** Werden für jedes Variablenpaar angezeigt. Das Kreuzprodukt der Abweichungen ist gleich der Summe der Produkte mittelwertkorrigierter Variablen. Dies ist der Zähler des Korrelationskoeffizienten nach Pearson. Die Kovarianz ist ein nicht standardisiertes Maß für den Zusammenhang zwischen zwei Variablen und ist gleich der Kreuzproduktabweichung dividiert durch *N*–1.

Fehlende Werte. Sie können eine der folgenden Optionen auswählen:

- **Paarweiser Fallausschluss.** Fälle mit fehlenden Werten für eine oder beide Variablen eines Paares für einen Korrelationskoeffizienten werden von der Analyse ausgeschlossen. Da jeder Koeffizient auf allen Fällen mit gültigen Codes für dieses bestimmte Variablenpaar basiert, werden in allen Berechnungen die maximal zugänglichen Informationen verwendet. Dies kann zu einer Menge von Koeffizienten führen, die auf einer variierenden Anzahl von Fällen basiert.
- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für Variablen werden von allen Korrelationen ausgeschlossen.

Zusätzliche Funktionen bei den Befehlen CORRELATIONS und NONPAR CORR

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Mit dem Unterbefehl `MATRIX` kann eine Korrelationsmatrix für Pearson-Korrelationen geschrieben werden. Diese kann anstelle von Rohdaten verwendet werden, um andere Analysen zu berechnen, beispielsweise die Faktorenanalyse.
- Mit dem Schlüsselwort `WITH` im Unterbefehl `VARIABLES` können die Korrelationen zwischen allen Variablen einer Liste und allen Variablen einer zweiten Liste berechnet werden.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
24

Partielle Korrelationen

Partielle Korrelationskoeffizienten beschreiben die Beziehung zwischen zwei Variablen. Die Prozedur "Partielle Korrelationen" berechnet diese Koeffizienten, wobei die Effekte von einer oder mehr zusätzlichen Variablen überprüft werden. Korrelationen sind Maße für lineare Zusammenhänge. Zwei Variablen können fehlerlos miteinander verbunden sein. Wenn es sich aber nicht um eine lineare Beziehung handelt, ist der Korrelationskoeffizient zur Messung des Zusammenhangs zwischen den beiden Variablen nicht geeignet.

Beispiel. Besteht eine Beziehung zwischen den Ausgaben für das Gesundheitswesen und den Krankheitsraten? Obwohl man annehmen könnte, eine solche Beziehung sei negativ, ergibt eine Studie eine signifikante *positive* Korrelation: mit ansteigenden Ausgaben im Gesundheitswesen scheinen die Krankheitsraten zuzunehmen. Durch die Kontrolle der Effekte aus der Häufigkeit der Besuche bei medizinischem Personal wird die beobachtete positive Korrelation praktisch eliminiert. Die Ausgaben im Gesundheitswesen und die Krankheitsraten scheinen lediglich in einer positiven Beziehung zu stehen, da mit steigender Finanzausstattung mehr Menschen Zugang zu medizinischer Versorgung haben, was zu mehr gemeldeten Krankheiten bei Ärzten und Krankenhäusern führt.

Statistiken. Für jede Variable: Anzahl der Fälle mit nichtfehlenden Werten, Mittelwert und Standardabweichung. Matrizen für partielle Korrelationen und Korrelationen nullter Ordnung mit Freiheitsgraden und Signifikanzniveaus.

Daten. Verwenden Sie symmetrische, quantitative Variablen.

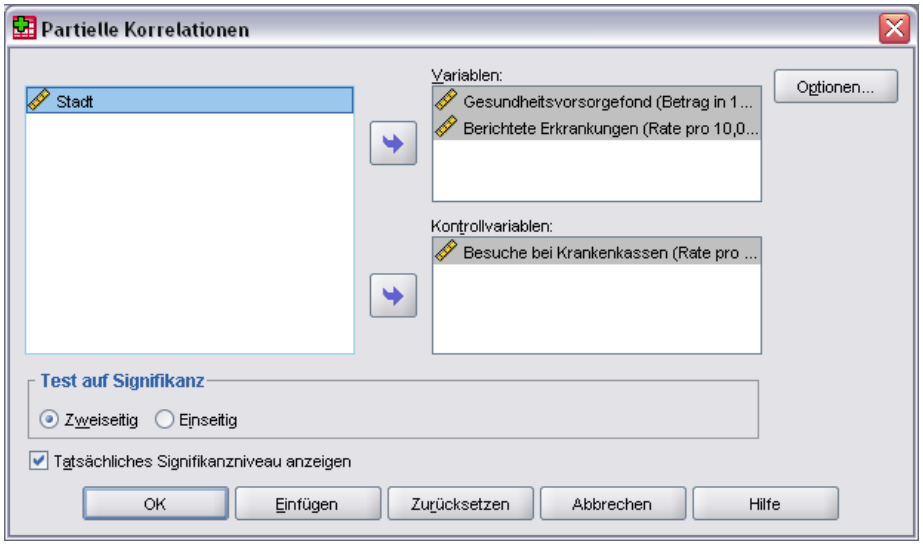
Annahmen. Die Prozedur "Partielle Korrelation" setzt für jedes Variablenpaar eine bivariate Normalverteilung voraus.

So lassen Sie partielle Korrelationen berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Korrelation
    Partiell

Abbildung 24-1
Dialogfeld "Partielle Korrelationen"



- Wählen Sie mindestens zwei numerische Variablen aus, für die partielle Korrelationen berechnet werden sollen.
- Wählen Sie mindestens eine numerische Kontroll-Variable aus.

Außerdem sind folgende Optionen verfügbar:

- **Test auf Signifikanz.** Sie können einseitige oder zweiseitige Wahrscheinlichkeiten wählen. Wenn Ihnen die Richtung des Zusammenhangs im voraus bekannt ist, wählen Sie Einseitig. Wählen Sie anderenfalls Zweiseitig.
- **Tatsächliches Signifikanzniveau anzeigen.** In der Standardeinstellung werden die Wahrscheinlichkeit sowie die Freiheitsgrade für jeden Korrelationskoeffizienten angezeigt. Wenn Sie diese Option deaktivieren, werden die Koeffizienten mit einem Signifikanzniveau von 0,05 mit einem Sternchen gekennzeichnet. Koeffizienten mit einem Signifikanzniveau von 0,01 werden mit einem doppelten Sternchen gekennzeichnet, und Freiheitsgrade werden unterdrückt. Diese Einstellung beeinflußt sowohl die Matrizen der partiellen Korrelationen als auch die der nullten Ordnung.

Partielle Korrelationen: Optionen

Abbildung 24-2
Dialogfeld "Partielle Korrelationen: Optionen"

Statistik. Sie können eine oder beide der folgenden Möglichkeiten auswählen:

- **Mittelwerte und Standardabweichungen.** Diese werden für jede Variable angezeigt. Außerdem wird die Anzahl der Fälle mit nichtfehlenden Werten angezeigt.
- **Korrelationen nullter Ordnung.** Hiermit wird eine einfache Matrix für Korrelationen zwischen allen Variablen (einschließlich Kontroll-Variablen) angezeigt.

Fehlende Werte. Sie können eine der folgenden Möglichkeiten wählen:

- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für Variablen (einschließlich Kontroll-Variablen) werden aus den Berechnungen ausgeschlossen.
- **Paarweiser Fallausschluss.** Bei der Berechnung der Korrelationen nullter Ordnung, die den partiellen Korrelationen zugrunde liegen, werden Fälle mit fehlenden Werten in einer oder beiden Variablen eines Variablenpaars nicht verwendet. Beim paarweisen Löschen wird der größtmögliche Teil der Daten verwendet. Die Anzahl der Fälle kann jedoch von Koeffizient zu Koeffizient variieren. Wenn das paarweise Löschen aktiviert ist, liegt den Freiheitsgraden eines bestimmten partiellen Koeffizienten die niedrigste Anzahl von Fällen zugrunde, die zur Berechnung einer der Korrelationen nullter Ordnung verwendet werden.

Zusätzliche Funktionen beim Befehl PARTIAL CORR

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Sie können eine Korrelationsmatrix nullter Ordnung einlesen und eine Matrix der partiellen Korrelationen schreiben (mit dem Unterbefehl `MATRIX`).
- Sie können partielle Korrelationen zwischen zwei Variablenlisten erstellen (mit dem Schlüsselwort `WITH` im Unterbefehl `VARIABLES`).
- Sie können mehrere Analysen berechnen lassen (mit mehren Unterbefehlen `VARIABLES`).
- Sie können die Ordnung für die Anfrage angeben (z. B. partielle Korrelationen sowohl erster als auch zweiter Ordnung), wenn Sie über zwei Kontrollvariablen verfügen (mit dem Unterbefehl `VARIABLES`).

- Sie können redundante Koeffizienten unterdrücken (mit dem Unterbefehl `FORMAT`).
- Sie können eine Matrix von einfachen Korrelationen anzeigen lassen, wenn einige Koeffizienten nicht berechnet werden können (mit dem Unterbefehl `STATISTICS`).

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
25

Distanzen

Durch diese Prozedur kann eine Vielzahl von Statistiken berechnet werden, indem Ähnlichkeiten oder Unähnlichkeiten (Distanzen) zwischen Paaren von Variablen oder Fällen gemessen werden. Diese Ähnlichkeits- oder Distanzmaße können dann bei anderen Prozeduren, beispielsweise der Faktorenanalyse, der Cluster-Analyse oder der multidimensionalen Skalierung zur Analyse komplexer Daten-Sets verwendet werden.

Beispiel. Ist es möglich, Ähnlichkeiten zwischen Paaren von Kraftfahrzeugen anhand bestimmter Merkmale zu messen, z. B. anhand des Hubraums, des Kraftstoffverbrauchs oder der Leistung? Durch die Berechnung von Ähnlichkeiten zwischen Kraftfahrzeugen können Sie besser einordnen, welche Fahrzeuge einander ähneln bzw. welche sich voneinander unterscheiden. Mit einer hierarchischen Cluster-Analyse oder einer multidimensionalen Skalierung auf die Ähnlichkeiten können Sie eine formale Analyse durchführen, um die zugrunde liegende Struktur zu untersuchen.

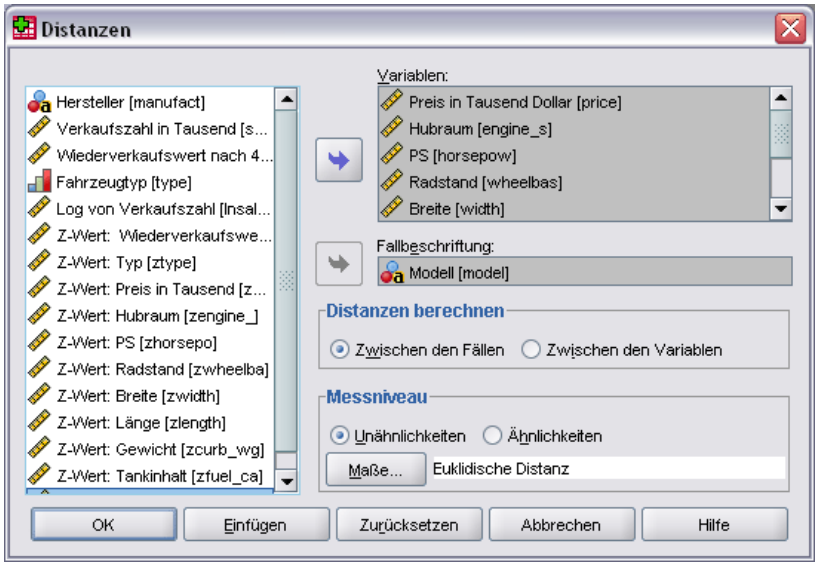
Statistiken. Unähnlichkeitsmaße (Distanzmaße) für Intervalldaten: Euklidischer Abstand, quadrierter Euklidischer Abstand, Tschebyscheff, Block, Minkowski oder ein benutzerdefiniertes Maß; für Häufigkeiten: Chi-Quadrat-Maß oder Phi-Quadrat-Maß; für Binärdaten: Euklidischer Abstand, quadrierter Euklidischer Abstand, Größendifferenz, Musterdifferenz, Varianz, Form und Distanzmaß nach Lance und Williams. Ähnlichkeitsmaße für Intervalldaten: Pearson-Korrelation oder Kosinus; für Binärdaten: Russel und Rao, einfache Übereinstimmung, Jaccard, Würfel-Ähnlichkeitsmaß, Ähnlichkeitsmaß nach Rogers und Tanimoto, Sokal und Sneath 1, Sokal und Sneath 2, Sokal und Sneath 3, Kulczynski 1, Kulczynski 2, Sokal und Sneath 4, Hamann, Lambda, Anderberg-*D*, Yule-*Y*, Yule-*Q*, Ochiai, Sokal und Sneath 5, Phi-4-Punkt-Korrelation oder Streuung.

So lassen Sie Distanzmatrizen berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Korrelation
    Distanzen...

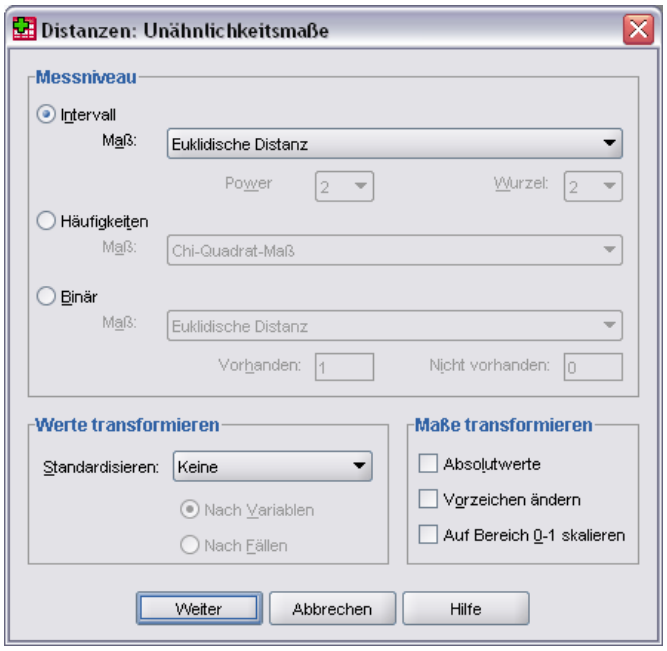
Abbildung 25-1
Dialogfeld "Distanzen"



- Wählen Sie mindestens eine numerische Variable zur Berechnung von Distanzen zwischen Fällen oder wählen Sie mindestens zwei numerische Variablen zur Berechnung von Distanzen zwischen Variablen.
- Wählen Sie im Gruppenfeld "Distanzen berechnen" eine andere Option aus, um Ähnlichkeiten zwischen Fällen oder Variablen zu berechnen.

Unähnlichkeitsmaße für Distanzen

Abbildung 25-2
Dialogfeld "Distanzen: Unähnlichkeitsmaße"



Wählen Sie aus dem Gruppenfeld "Maß" die Option aus, die Ihrem Datentyp entspricht ("Intervall", "Häufigkeiten" oder "Binär"). Wählen Sie dann aus dem Dropdown-Listenfeld ein Maß aus, das diesem Datentyp entspricht. Die folgenden Maße sind je nach Datentyp verfügbar:

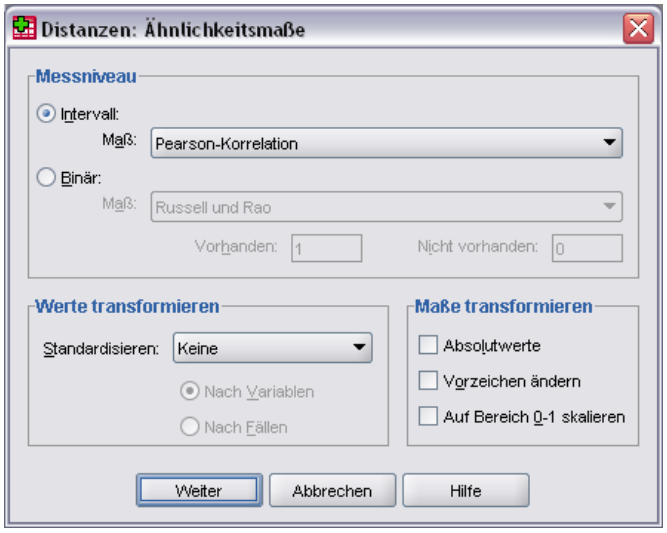
- **Intervall.** Euklidischer Abstand, quadrierter Euklidischer Abstand, Tschebyscheff, Block, Minkowski oder ein benutzerdefiniertes Maß.
- **Häufigkeiten.** Chi-Quadrat-Maß oder Phi-Quadrat-Maß.
- **Binär.** Euklidischer Abstand, quadrierter Euklidischer Abstand, Größendifferenz, Musterdifferenz, Varianz, Form und Distanzmaß nach Lance und Williams. (Geben Sie Werte in die Felder "Vorhanden" und "Nicht vorhanden" ein, um anzugeben, welche beiden Werte sinnvoll sind; alle übrigen Werte werden durch die Distanzmaße ignoriert.)

Im Gruppenfeld "Werte transformieren" können Sie festlegen, ob die Datenwerte für Fälle oder Werte *vor* dem Berechnen von Ähnlichkeiten für Fälle oder Variablen standardisiert werden. Diese Transformationen sind nicht auf binäre Daten anwendbar. Die verfügbaren Standardisierungsmethoden sind "Z-Scores", "Bereich –1 bis 1", "Bereich 0 bis 1", "Maximale Größe von 1", "Mittelwert 1" und "Standardabweichung 1".

Im Gruppenfeld "Maße transformieren" können Sie festlegen, ob die durch das Distanzmaß erzeugten Werte transformiert werden. Dies erfolgt, nachdem das Distanzmaß berechnet wurde. Zu den verfügbaren Optionen zählen Absolutwerte, Ändern des Vorzeichens und Skalieren auf den Bereich 0–1.

Ähnlichkeitsmaße für Distanzen

Abbildung 25-3
Dialogfeld "Distanzen: Ähnlichkeitsmaße"



Wählen Sie aus dem Gruppenfeld "Maß" die Option aus, die Ihrem Datentyp entspricht ("Intervall" oder "Binär"). Wählen Sie dann aus dem Dropdown-Listenfeld ein Maß aus, das diesem Datentyp entspricht. Die folgenden Maße sind je nach Datentyp verfügbar:

- **Intervall.** Pearson-Korrelation oder Kosinus
- **Binär.** Russel und Rao, einfache Übereinstimmung, Jaccard, Würfel-Ähnlichkeitsmaß, Ähnlichkeitsmaß nach Rogers und Tanimoto, Ähnlichkeitsmaße nach Sokal und Sneath 1 bis 5, Kulczynski 1, Kulczynski 2, Sokal und Sneath 4, Hamann, Lambda, Anderberg-*D*, Yule-*Y*, Yule-*Q*, Ochiai, Sokal und Sneath 5, Phi-4-Punkt-Korrelation oder Streuung. (Geben Sie Werte in die Felder "Vorhanden" und "Nicht vorhanden" ein, um anzugeben, welche beiden Werte sinnvoll sind; alle übrigen Werte werden durch die Distanzmaße ignoriert.)

Im Gruppenfeld "Werte transformieren" können Sie festlegen, ob die Datenwerte für Fälle oder Variablen vor dem Berechnen von Ähnlichkeiten standardisiert werden. Diese Transformationen sind nicht auf binäre Daten anwendbar. Die verfügbaren Standardisierungsmethoden sind "*Z*-Scores", "Bereich –1 bis 1", "Bereich 0 bis 1", "Maximale Größe von 1", "Mittelwert 1" und "Standardabweichung 1".

Im Gruppenfeld "Maße transformieren" können Sie festlegen, ob die durch das Distanzmaß erzeugten Werte transformiert werden. Dies erfolgt, nachdem das Distanzmaß berechnet wurde. Zu den verfügbaren Optionen zählen Absolutwerte, Ändern des Vorzeichens und Skalieren auf den Bereich 0–1.

Zusätzliche Funktionen beim Befehl PROXIMITIES

In der Prozedur “Distanzen” wird die Befehlssyntax von `PROXIMITIES` verwendet. Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Angeben einer ganze Zahl als Exponent für das Minkowski-Distanzmaß
- Angeben von beliebigen Ganzzahlen als Exponent und Wurzel für ein benutzerdefiniertes Distanzmaß

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
26

Lineare Regression

Mit “Lineare Regression” werden die Koeffizienten der linearen Gleichung unter Einbeziehung einer oder mehrerer unabhängiger Variablen geschätzt, die den Wert der abhängigen Variablen am besten vorhersagen. Sie können beispielsweise den Versuch unternehmen, die Jahresverkaufsbilanz eines Verkäufers (die abhängige Variable) nach unabhängigen Variablen wie Alter, Bildungsstand und Anzahl der Berufsjahre vorherzusagen.

Beispiel. Besteht ein Zusammenhang zwischen der Anzahl der in einer Saison gewonnenen Spiele eines Basketball-Teams und der pro Spiel erzielten mittleren Punktezahl des Teams? Einem Streudiagramm läßt sich entnehmen, dass zwischen diesen Variablen eine lineare Beziehung besteht. Die Anzahl gewonnener Spiele und die erzielte Punktezahl des Gegners stehen gleichfalls in linearer Beziehung zueinander. Diese Variablen enthalten eine negative Beziehung. Einer steigenden Anzahl gewonnener Spiele steht eine fallende mittlere Punktezahl des Gegners gegenüber. Mit der linearen Regression können Sie die Beziehung dieser Variablen modellieren. Mit einem geeigneten Modell lassen sich Spielgewinne von Teams vorhersagen.

Statistiken. Für jede Variable: Anzahl gültiger Fälle, Mittelwert und Standardabweichung. Für jedes Modell: Regressionskoeffizienten, Korrelationsmatrix, Teil- und partielle Korrelationen, multiples R, R^2, korrigiertes R^2, Änderung in R^2, Standardfehler der Schätzung, Tabelle der Varianzanalyse, vorhergesagte Werte und Residuen. Außerdem 95%-Konfidenzintervalle für jeden Regressionskoeffizienten, Varianz-Kovarianz-Matrix, Inflationsfaktor der Varianz, Toleranz, Durbin-Watson-Test, Distanzmaße (Mahalanobis, Cook und Hebelwerte), DfBeta, DfFit, Vorhersageintervalle und fallweise Diagnose. Diagramme: Streudiagramme, partielle Diagramme, Histogramme und Normalverteilungsdiagramme.

Daten. Die abhängigen und die unabhängigen Variablen müssen quantitativ sein. Kategoriale Variablen, wie beispielsweise Religion, Studienrichtung oder Wohnsitz, müssen in binäre (Dummy-)Variablen oder andere Typen von Kontrast-Variablen umkodiert werden.

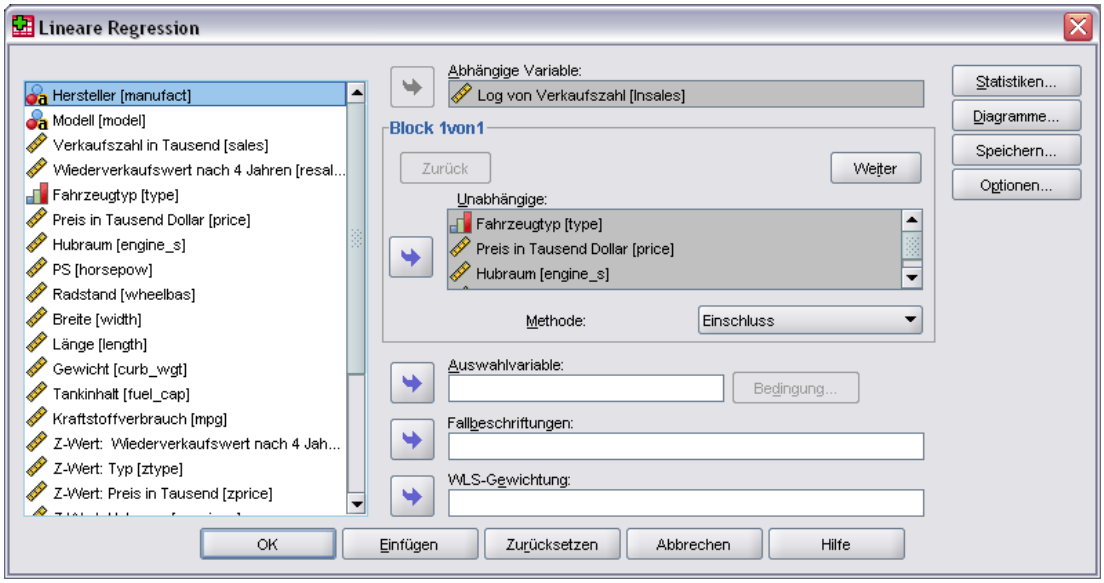
Annahmen. Für jeden Wert der unabhängigen Variablen muss die abhängige Variable normalverteilt vorliegen. Die Varianz der Verteilung der abhängigen Variablen muss für alle Werte der unabhängigen Variablen konstant sein. Die Beziehung zwischen der abhängigen Variablen und allen unabhängigen Variablen sollte linear sein, und alle Beobachtungen sollten unabhängig sein.

So lassen Sie eine lineare Regressionsanalyse berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Regression
    Linear...

Abbildung 26-1
Dialogfeld "Lineare Regression"



- Wählen Sie im Dialogfeld "Lineare Regression" eine numerische abhängige Variable aus.
- Wählen Sie eine oder mehrere numerische unabhängige Variablen aus.

Die folgenden Optionen sind verfügbar:

- Unabhängige Variablen können in Blöcken zusammengefaßt werden, und es können verschiedene Einschlussmethoden für unterschiedliche Untergruppen von Variablen angegeben werden.
- Auswahlvariablen zum Begrenzen der Analyse auf eine Untergruppe von Fällen mit einem bestimmten Wert oder bestimmten Werten für diese Variable können ausgewählt werden.
- Es können Variablen zur Fallunterscheidung ausgewählt werden, um Punkte in Diagrammen zu identifizieren.
- Wählen Sie eine numerische Variable für die WLS-Gewichtung aus, um eine Analyse der gewichteten kleinsten Quadrate durchzuführen.

WLS (Gewichtete kleinste Quadrate). Hiermit können Sie ein Modell gewichteter kleinster Quadrate berechnen. Die Datenpunkte werden mit dem reziproken Wert ihrer Varianzen gewichtet. Dies bedeutet, dass Beobachtungen mit großen Varianzen die Analyse weniger beeinflussen als Beobachtungen mit kleinen Varianzen. Wenn der Wert der Gewichtungsvariablen null, negativ oder fehlend ist, wird der Fall aus der Analyse ausgeschlossen.

Lineare Regression: Methode zur Auswahl von Variablen

Durch die Auswahl der Methode können Sie festlegen, wie unabhängige Variablen in die Analyse eingeschlossen werden. Anhand verschiedener Methoden können Sie eine Vielfalt von Regressionsmodellen mit demselben Satz von Variablen erstellen.

- **Einschluss (Regression).** Eine Prozedur für die Variablenauswahl, bei der alle Variablen eines Blocks in einem einzigen Schritt aufgenommen werden.
- **Schrittweise.** Bei jedem Schritt wird die noch nicht in der Gleichung enthaltene unabhängige Variable mit der kleinsten F-Wahrscheinlichkeit aufgenommen, sofern diese Wahrscheinlichkeit klein genug ist. Bereits in der Regressionsgleichung enthaltene Variablen werden entfernt, wenn ihre F-Wahrscheinlichkeit hinreichend groß wird. Das Verfahren endet, wenn keine Variablen mehr für Aufnahme oder Ausschluss infrage kommen.
- **Entfernen.** Ein Verfahren zur Variablenauswahl, bei dem alle Variablen eines Blocks in einem Schritt ausgeschlossen werden.
- **Rückwärtselimination.** Eine Methode zur Variablenauswahl, bei der alle Variablen in die Gleichung aufgenommen und anschließend sequenziell ausgeschlossen werden. Die Variable mit der kleinsten Teilkorrelation zur abhängigen Variablen wird als erste für den Ausschluss in Betracht gezogen. Wenn sie das Ausschlusskriterium erfüllt, wird sie entfernt. Nach dem Ausschluss der ersten Variablen wird die nächste Variable mit der kleinsten Teilkorrelation in Betracht gezogen. Das Verfahren wird beendet, wenn keine Variablen mehr zur Verfügung stehen, die die Ausschlusskriterien erfüllen.
- **Vorwärtsselektion.** Ein Verfahren zur schrittweisen Variablenauswahl, in dem die Variablen nacheinander in das Modell aufgenommen werden. Die erste Variable, die in Betracht gezogen wird, ist die mit der größten positiven bzw. negativen Korrelation mit der abhängigen Variablen. Diese Variable wird nur dann in die Gleichung aufgenommen, wenn sie das Aufnahmekriterium erfüllt. Wenn die erste Variable aufgenommen wurde, wird als Nächstes die unabhängige Variable mit der größten partiellen Korrelation betrachtet. Das Verfahren endet, wenn keine verbliebene Variable das Aufnahmekriterium erfüllt.

Die Signifikanzwerte in Ihrer Ausgabe basieren auf der Berechnung eines einzigen Modells. Deshalb sind diese generell ungültig, wenn eine schrittweise Methode (schrittweise, vorwärts oder rückwärts) verwendet wird.

Alle Variablen müssen das Toleranzkriterium erfüllen, um unabhängig von der angegebenen Einschlussmethode in die Gleichung einbezogen zu werden. In der Standardeinstellung beträgt der Toleranzwert 0,0001. Eine Variable wird auch dann nicht eingeschlossen, wenn dadurch die Toleranz einer Variablen im Modell unter das Toleranzkriterium abfallen würde.

Alle ausgewählten unabhängigen Variablen werden einem einzigen Regressionsmodell hinzugefügt. Sie können jedoch verschiedene Einschlussmethoden für unterschiedliche Untergruppen von Variablen angeben. Beispielsweise können Sie einen Block von Variablen durch schrittweises Auswählen und einen zweiten Block durch Vorwärtsselektion in das Regressionsmodell einschließen. Um einem Regressionsmodell einen zweiten Block von Variablen hinzuzufügen, klicken Sie auf Weiter.

Lineare Regression: Bedingung aufstellen

Abbildung 26-2
Dialogfeld "Lineare Regression: Bedingung aufstellen"

Die durch die Auswahlbedingung definierten Fälle werden in die Analyse eingeschlossen. Wenn Sie für die Variable beispielsweise gleich wählen und als Wert 5 eingeben, werden nur Fälle in die Analyse einbezogen, für die der Wert der gewählten Variablen gleich 5 ist. Ein String-Wert ist ebenfalls möglich.

Lineare Regression: Diagramme

Abbildung 26-3
Dialogfeld "Lineare Regression: Diagramme"

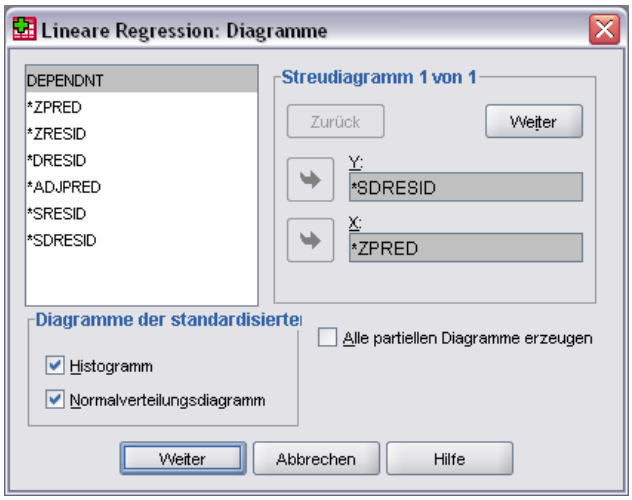


Diagramme können beim Validieren der Annahmen von Normalverteilung, Linearität und Varianz-Gleichheit hilfreich sein. Diagramme dienen auch zum Auffinden von Ausreißern, ungewöhnlichen Beobachtungen und Einflußfällen. Nachdem sie als neue Variablen gespeichert wurden, stehen im Daten-Editor vorhergesagte Werte, Residuen und andere diagnostische Hilfsmittel zum Erstellen von Diagrammen mit den unabhängigen Variablen zur Verfügung. Folgende Diagramme sind verfügbar:

Streudiagramme. Sie können je zwei der folgenden Elemente auftragen: die abhängige Variable, standardisierte vorhergesagte Werte, standardisierte Residuen, ausgeschlossene Residuen, korrigierte vorhergesagte Werte, studentisierte Residuen oder studentisierte ausgeschlossene Residuen. Tragen Sie die standardisierten Residuen über den standardisierten vorhergesagten Werten auf, um auf Linearität und Varianz-Gleichheit zu überprüfen.

Liste der Quellvariablen. Listet die abhängigen Variablen (DEPENDNT) und die folgenden vorhergesagten Variablen und Residuen-Variablen auf: standardisierte vorhergesagte Werte (*ZPRED), standardisierte Residuen (*ZRESID), ausgeschlossene Residuen (*DRESID), korrigierte vorhergesagte Werte (*ADJPRED), studentisierte Residuen (*SRESID) und studentisierte ausgeschlossene Residuen (*SDRESID).

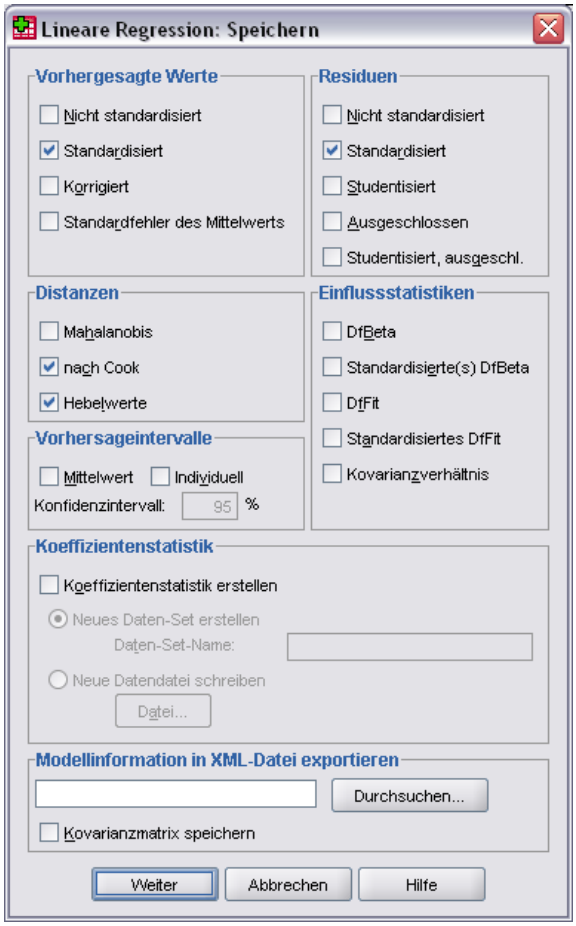
Alle partiellen Diagramme erzeugen. Erzeugt Streudiagramme der Residuen aller unabhängigen Variablen und der Residuen der abhängigen Variablen, wenn für den Rest der unabhängigen Variablen beide Variablen einer getrennten Regression unterzogen werden. Zum Erzeugen eines partiellen Diagramms müssen mindestens zwei unabhängige Variablen in der Gleichung enthalten sein.

Diagramme der standardisierten Residuen. Sie können Histogramme standardisierter Residuen und Normalverteilungsdiagramme anfordern, welche die Verteilung standardisierter Residuen mit einer Normalverteilung vergleichen.

Beim Anfordern von Diagrammen werden Auswertungsstatistiken für standardisierte vorhergesagte Werte und standardisierte Residuen (*ZPRED und *ZRESID) angezeigt.

Lineare Regression: Speichern von neuen Variablen

Abbildung 26-4
Dialogfeld "Lineare Regression: Speichern"



Vorhergesagte Werte, Residuen und andere für die Diagnose nützliche Statistiken können gespeichert werden. Mit jedem Auswahlvorgang werden Ihrer Datendatei eine oder mehrere neue Variablen hinzugefügt.

Vorhergesagte Werte. Dies sind die nach dem Regressionsmodell für jeden Fall vorhergesagten Werte.

- **Nicht standardisiert.** Der Wert, den das Modell für die abhängige Variable vorhersagt.
- **Standardisiert.** Eine Transformation jedes vorhergesagten Werts in dessen standardisierte Form. Das heißt, dass die Differenz zwischen dem vorhergesagten Wert und dem mittleren vorhergesagten Wert durch die Standardabweichung der vorhergesagten Werte geteilt wird. Standardisierte vorhergesagte Werte haben einen Mittelwert von 0 und eine Standardabweichung von 1.

- **Korrigiert.** Der vorhergesagte Wert für einen Fall, wenn dieser Fall von der Berechnung der Regressionskoeffizienten ausgeschlossen ist.
- **Standardfehler des Mittelwerts.** Standardfehler der vorhergesagten Werte. Ein Schätzer der Standardabweichung des Durchschnittswerts der abhängigen Variablen für die Fälle, die dieselben Werte für die unabhängigen Variablen haben.

Distanzen. Dies sind Maße zum Auffinden von Fällen mit ungewöhnlichen Wertekombinationen bei der unabhängigen Variablen und von Fällen, die einen großen Einfluß auf das Modell haben könnten.

- **Mahalanobis.** Dieses Maß gibt an, wie weit die Werte der unabhängigen Variablen eines Falles vom Mittelwert aller Fälle abweichen. Ein großer Mahalanobis-Abstand charakterisiert einen Fall, der bei einer oder mehreren unabhängigen Variablen Extremwerte besitzt.
- **Cook.** Ein Maß dafür, wie stark sich die Residuen aller Fälle ändern würden, wenn ein spezieller Fall von der Berechnung der Regressionskoeffizienten ausgeschlossen würde. Ein großer Wert der Cook-Distanz zeigt an, dass der Ausschluss eines Falles von der Berechnung der Regressionskoeffizienten die Koeffizienten substanziell verändert.
- **Hebelwerte.** Werte, die den Einfluss eines Punktes auf die Anpassung der Regression messen. Der zentrierte Wert für die Hebelwirkung bewegt sich zwischen 0 (kein Einfluss auf die Anpassung) und (N-1)/N.

Vorhersageintervalle. Die oberen und unteren Grenzen sowohl für Mittelwert als auch für einzelne Vorhersageintervalle.

- **Mittelwert.** Unter- und Obergrenze (zwei Variablen) für das Vorhersageintervall für den mittleren vorhergesagten Wert.
- **Individuell.** Unter- und Obergrenze (zwei Variablen) für das Vorhersageintervall der abhängigen Variablen für einen Einzelfall.
- **Konfidenzintervall.** Geben Sie einen Wert zwischen 1 und 99,99 ein, um das Konfidenzniveau für die beiden Vorhersageintervalle festzulegen. Wählen Sie “Mittelwert” oder “Individuell” aus, bevor Sie diesen Wert eingeben. Typische Werte für Konfidenzniveaus sind 90, 95 und 99.

Residuen. Der tatsächliche Wert der abhängigen Variablen minus des vorhergesagten Werts aus der Regressionsgleichung.

- **Nicht standardisiert.** Die Differenz zwischen einem beobachteten Wert und dem durch das Modell vorhergesagten Wert.
- **Standardisiert.** Der Quotient aus dem Residuum und einem Schätzer seiner Standardabweichung. Standardisierte Residuen, auch bekannt als Pearson-Residuen, haben einen Mittelwert von 0 und eine Standardabweichung von 1.
- **Studentisiert.** Ein Residuum, das durch seine geschätzte Standardabweichung geteilt wird, die je nach der Distanz zwischen den Werten der unabhängigen Variablen des Falles und dem Mittelwert der unabhängigen Variablen von Fall zu Fall variiert.

- **Ausgeschlossen.** Das Residuum für einen Fall, wenn dieser Fall nicht in die Berechnung der Regressionskoeffizienten eingegangen ist. Es ist die Differenz zwischen dem Wert der abhängigen Variablen und dem korrigierten Schätzwert.
- **Studentisiert, ausgeschl..** Der Quotient aus dem ausgeschlossenen Residuum eines Falles und seinem Standardfehler. Die Differenz zwischen einem studentisierten ausgeschlossenen Residuum und dem zugehörigen studentisierten Residuum gibt an, welchen Unterschied die Entfernung eines Falles für dessen eigene Vorhersage bewirkt.

Einflußstatistiken. Die Änderung in den Regressionskoeffizienten (DfBeta[s]) und vorhergesagten Werten (DfFit), die sich aus dem Ausschluss eines bestimmten Falls ergibt. Standardisierte DfBetas- und DfFit-Werte stehen zusammen mit dem Kovarianzverhältnis zur Verfügung.

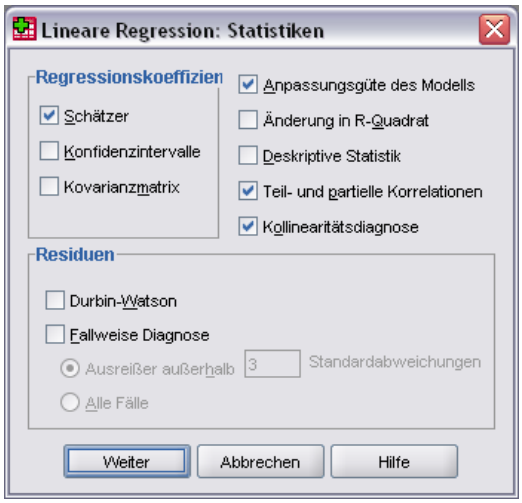
- **Differenz in Beta (DfBeta(s)).** Die Differenz im Beta-Wert entspricht der Änderung im Regressionskoeffizienten, die sich aus dem Ausschluss eines bestimmten Falls ergibt. Für jeden Term im Modell, einschließlich der Konstanten, wird ein Wert berechnet.
- **Standardisiertes DfBeta.** Die standardisierte Differenz im Beta-Wert. Die Änderung des Regressionskoeffizienten, die sich durch den Ausschluss eines bestimmten Falls ergibt. Es empfiehlt sich, Fälle mit absoluten Werten größer als 2 geteilt durch die Quadratwurzel von N zu überprüfen, wenn N die Anzahl der Fälle darstellt. Für jeden Term im Modell, einschließlich der Konstanten, wird ein Wert berechnet.
- **Differenz im vorhergesagten Wert (DfFit).** Die Änderung im vorhergesagten Wert, die sich aus dem Ausschluss eines bestimmten Falls ergibt.
- **Standardisiertes DfFit.** Die standardisierte Differenz im Anpassungswert. Die Änderung des vorhergesagten Werts, die sich durch den Ausschluss eines bestimmten Falls ergibt. Es empfiehlt sich, Fälle mit absoluten Werten größer als 2 geteilt durch die Quadratwurzel von p/N zu überprüfen, wobei p die Anzahl der unabhängigen Variablen im Modell und N die Anzahl der Fälle darstellt.
- **Kovarianzverhältnis.** Das Verhältnis der Determinante der Kovarianzmatrix bei Ausschluss eines bestimmten Falles von der Berechnung des Regressionskoeffizienten zur Determinante der Kovarianzmatrix bei Einschluss aller Fälle. Wenn der Quotient dicht bei 1 liegt, beeinflusst der ausgeschlossene Fall die Kovarianzmatrix nur unwesentlich.

Koeffizientenstatistik. Speichert den Regressionskoeffizienten in einem Daten-Set oder in einer Datendatei. Daten-Sets sind für die anschließende Verwendung in der gleichen Sitzung verfügbar, werden jedoch nicht als Dateien gespeichert, sofern Sie diese nicht ausdrücklich vor dem Beenden der Sitzung speichern. Die Namen von Daten-Sets müssen den Regeln zum Benennen von Variablen entsprechen. Für weitere Informationen siehe Variablennamen in Kapitel 5 auf S. 83.

Modellinformation in XML-Datei exportieren. Parameterschätzer und (wahlweise) ihre Kovarianzen werden in die angegebene Datei exportiert. SmartScore und SPSS Server (gesondertes Produkt) können anhand dieser Modelldatei die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Lineare Regression: Statistiken

Abbildung 26-5
Dialogfeld "Statistiken"



Folgende Statistiken sind verfügbar:

Regressionskoeffizienten. Mit Schätzer zeigen Sie den Regressionskoeffizienten *B*, den Standardfehler von *B*, das Beta des standardisierten Koeffizienten, den *t*-Wert für *B* und das zweiseitige Signifikanzniveau von *t* an. Mit Konfidenzintervalle werden die 95%-Konfidenzintervalle für jeden Regressionskoeffizienten oder eine Kovarianzmatrix angezeigt. Mit Kovarianzmatrix wird eine Varianz-Kovarianz-Matrix von Regressionskoeffizienten mit Kovarianzen angezeigt, die nicht auf der Diagonalen liegen, und Varianzen, die auf der Diagonalen liegen. Außerdem wird eine Korrelationsmatrix angezeigt.

Anpassungsgüte des Modells. Die aufgenommenen und entfernten Variablen aus dem Modell werden aufgelistet, und die folgenden Statistiken der Anpassungsgüte werden angezeigt: multiples *R*, R^2 und korrigiertes R^2, Standardfehler der Differenz und eine Tabelle zur Varianzanalyse.

Änderung in R-Quadrat. Die Änderung in R^2, die aus dem Hinzufügen oder Entfernen einer unabhängigen Variablen resultiert. Wenn die durch eine Variable bewirkte Änderung in R^2 groß ist, bedeutet dies, dass diese Variable eine aussagekräftige Einflußvariable für die abhängige Variable ist.

Deskriptive Statistiken. Liefert die Anzahl gültiger Fälle, Mittelwert und Standardabweichung für jede Variable in der Analyse. Außerdem werden eine Korrelationsmatrix mit einem einseitigen Signifikanzniveau und die Anzahl der Fälle für jede Korrelation angezeigt.

Partielle Korrelation. Die Korrelation, die zwischen zwei Variablen verbleibt, nachdem die Korrelation entfernt wurde, die aus dem wechselseitigen Zusammenhang mit den anderen Variablen stammt. Die Korrelation zwischen der abhängigen Variablen und einer unabhängigen Variablen, wenn die linearen Effekte der anderen unabhängigen Variablen im Modell aus der unabhängigen Variablen entfernt wurden.

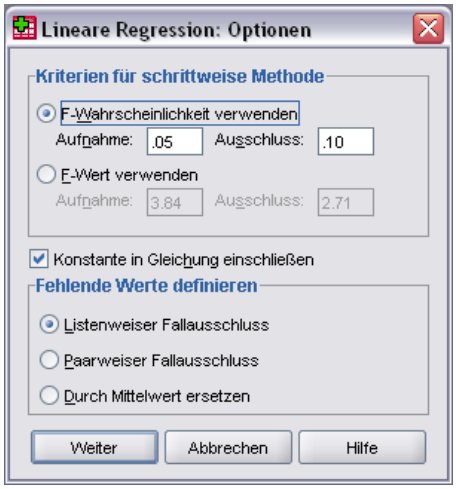
Teilkorrelation. Die Korrelation zwischen der abhängigen Variablen und einer unabhängigen Variablen, wenn die linearen Effekte der anderen unabhängigen Variablen im Modell aus der unabhängigen Variablen entfernt wurden. Die Korrelation entspricht der Änderung in R-Quadrat beim Addieren einer Variablen zu einer Gleichung. Zuweilen als semipartielle Korrelation bezeichnet.

Kollinearitätsdiagnose. Kollinearität (oder Multikollinearität) ist die unerwünschte Situation, in der eine unabhängige Variable eine lineare Funktion anderer unabhängiger Variablen ist. Eigenwerte der skalierten und unzentrierten Kreuzproduktmatrix, Bedingungsindexe und Proportionen der Varianzzerlegung werden zusammen mit Varianzfaktoren (VIF) und Toleranzen für einzelne Variablen angezeigt.

Residuen. Hiermit werden der Durbin-Watson-Test für Reihenkorrelationen der Residuen sowie die fallweise Diagnose für die Fälle angezeigt, die das Auswahlkriterium (Ausreißer über *n* Standardabweichungen) erfüllen.

Lineare Regression: Optionen

Abbildung 26-6
Dialogfeld "Lineare Regression: Optionen"



Die folgenden Optionen sind verfügbar:

Kriterien für schrittweise Methode. Diese Optionen eignen sich für den Fall, dass die Vorwärts-, Rückwärts- oder schrittweise Methode der Variablenauswahl angegeben wurde. Variablen im Modell können abhängig entweder von der Signifikanz (Wahrscheinlichkeit) des *F*-Werts oder vom *F*-Wert selbst eingeschlossen oder entfernt werden.

- **F-Wahrscheinlichkeit verwenden.** Eine Variable wird in das Modell aufgenommen, wenn das Signifikanzniveau ihres F-Werts kleiner ist als der Aufnahmewert. Sie wird ausgeschlossen, wenn das Signifikanzniveau größer ist als der Ausschlusswert. Der Aufnahmewert muss kleiner sein als der Ausschlusswert und beide Werte müssen positiv sein. Um mehr Variablen

in das Modell aufzunehmen, erhöhen Sie den Aufnahmewert. Um mehr Variablen aus dem Modell auszuschließen, senken Sie den Ausschlusswert.

- **F-Wert verwenden.** Eine Variable wird in ein Modell aufgenommen, wenn ihr F-Wert größer ist als der Aufnahmewert. Sie wird ausgeschlossen, wenn der F-Wert kleiner ist als der Ausschlusswert. Der Aufnahmewert muss größer sein als der Ausschlusswert und beide Werte müssen positiv sein. Um mehr Variablen in das Modell aufzunehmen, senken Sie den Aufnahmewert. Um mehr Variablen aus dem Modell auszuschließen, erhöhen Sie den Ausschlusswert.

Konstante in Gleichung einschließen. Als Voreinstellung enthält das Regressionsmodell einen konstanten Term. Wenn diese Option deaktiviert ist, wird die Regression durch den Ursprung gezwungen (selten verwendet). Manche Resultate einer durch den Ursprung verlaufenden Regression lassen sich nicht mit denen einer Regression vergleichen, die eine Konstante aufweist. Beispielsweise kann R^2 nicht in der üblichen Weise interpretiert werden.

Fehlende Werte. Sie können eine der folgenden Optionen auswählen:

- **Listenweiser Fallausschluss.** Nur Fälle mit gültigen Werten für alle Variablen werden in die Analyse einbezogen.
- **Paarweiser Fallausschluss.** Fälle mit vollständigen Daten für das korrelierte Variablenpaar werden zum Berechnen des Korrelationskoeffizienten verwendet, auf dem die Regressionsanalyse basiert. Freiheitsgrade basieren auf dem minimalen paarweisen *N*.
- **Durch Mittelwert ersetzen.** Alle Fälle werden für Berechnungen verwendet, wobei der Mittelwert der Variablen die fehlenden Beobachtungen ersetzt.

Zusätzliche Funktionen beim Befehl REGRESSION

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Schreiben einer Korrelationsmatrix oder Einlesen einer Matrix anstelle der Rohdaten, um eine Regressionsanalyse zu erhalten (mit dem Unterbefehl `MATRIX`)
- Angeben von Toleranzniveaus (mit dem Unterbefehl `CRITERIA`)
- Berechnen mehrerer Modelle für dieselben oder unterschiedliche abhängige Variablen (mit den Unterbefehlen `METHOD` und `DEPENDENT`)
- Berechnen zusätzlicher Statistiken (mit den Unterbefehlen `DESCRIPTIVES` und `STATISTICS`)

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
27

Ordinale Regression

Die ordinale Regression ermöglicht es, die Abhängigkeit einer polytomen ordinalen Antwortvariablen von einer Gruppe von Einflußvariablen zu modellieren. Bei diesen kann es sich um Faktoren oder Kovariaten handeln. Die Gestaltung der ordinalen Regression basiert auf der Methodologie von McCullagh (1980, 1998). In der Syntax wird diese Prozedur als `PLUM` bezeichnet.

Das Standardverfahren der linearen Regressionsanalyse beinhaltet die Minimierung der Summe von quadrierten Differenzen zwischen einer Antwortvariablen (abhängig) und einer gewichteten Kombination von Einflußvariablen (unabhängig). Die geschätzten Koeffizienten geben die Auswirkung einer Änderung in den Einflußvariablen auf die Antwortvariable wieder. Es wird angenommen, daß die Antwortvariable in dem Sinne numerisch ist, daß die Änderungen im Niveau der Antwortvariablen über die gesamte Spannweite der Antwortvariablen gleich sind. So beträgt die Differenz in der Körpergröße zwischen einer Person mit einer Größe von 150 cm und einer Person mit einer Größe von 140 cm beispielsweise 10 cm. Diese Angabe hat die gleiche Bedeutung wie die Differenz zwischen einer Person mit einer Größe von 210 cm und einer Person mit einer Größe von 200 cm. Bei ordinalen Variablen sind diese Beziehungen jedoch nicht notwendigerweise gegeben. Bei diesen Variablen kann die Auswahl und Anzahl von Antwortkategorien willkürlich ausfallen.

Beispiel. Die ordinale Regression kann verwendet werden, um die Reaktion von Patienten auf verschiedene Dosierungen eines Medikaments zu untersuchen. Die möglichen Reaktionen werden als *keine*, *mild*, *moderat* bzw. *stark* kategorisiert. Der Unterschied zwischen einer milden und einer moderaten Reaktion kann schwer oder gar nicht quantifiziert werden. Er gründet sich vielmehr auf reine Wahrnehmung. Der Unterschied zwischen einer milden und einer moderaten Reaktion kann darüber hinaus auch größer oder kleiner als der Unterschied zwischen einer moderaten und einer starken Reaktion ausfallen.

Statistiken und Diagramme. Beobachtete und erwartete Häufigkeiten und kumulative Häufigkeiten, Pearson-Residuen für Häufigkeiten und kumulative Häufigkeiten, beobachtete und erwartete Wahrscheinlichkeiten, beobachtete und erwartete kumulative Wahrscheinlichkeiten jeder Antwortkategorie nach Kovariaten-Struktur, asymptotische Korrelations- und Kovarianzmatrizen der Parameterschätzer, Pearson-Chi-Quadrat und Likelihood-Quotienten-Chi-Quadrat, Statistik der Anpassungsgüte, Iterationsprotokoll, Test der Annahme von parallelen Linien, Parameterschätzer, Standardfehler, Konfidenzintervalle sowie R^2 nach Cox und Snell, Nagelkerke und McFadden.

Daten. Es wird angenommen, dass die abhängige Variable ordinal ist. Sie kann eine numerische oder eine String-Variable sein. Die Reihenfolge richtet sich nach einer aufsteigenden Sortierung der Werte der abhängigen Variablen. Der niedrigste Wert entspricht der ersten Kategorie. Es wird angenommen, daß die Faktorvariablen kategorial sind. Die Kovariaten-Variablen müssen

numerisch sein. Beachten Sie, daß die Verwendung von mehr als einer stetigen Kovariate leicht zu einer sehr umfangreichen Tabelle mit Zellen-Wahrscheinlichkeiten führen kann.

Annahmen. Es darf nur eine Responsevariable vorhanden sein, und diese muß angegeben werden. Zusätzlich wird angenommen, daß die Antworten bei jeder eindeutigen Wertstruktur in den unabhängigen Variablen unabhängige multinomiale Variablen darstellen.

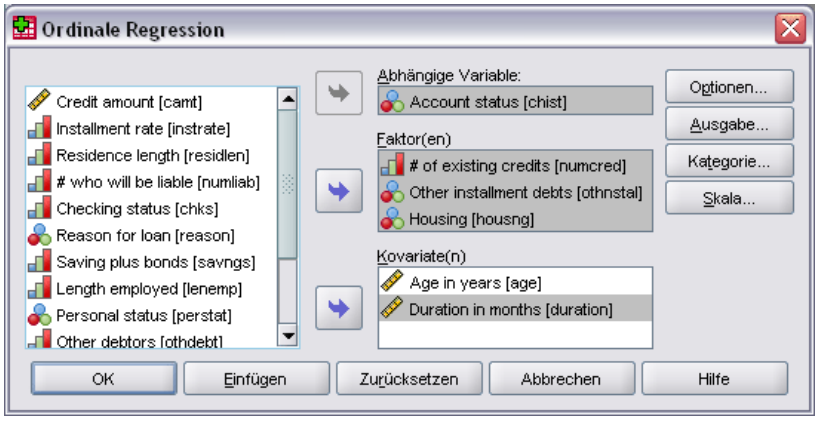
Verwandte Prozeduren. Bei der nominalen logistischen Regression werden ähnliche Modelle für nominale abhängige Variablen verwendet.

Berechnen einer ordinalen Regression

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
Regression
Ordinal...

Abbildung 27-1
Dialogfeld "Ordinale Regression"



▶ Wählen Sie eine abhängige Variable aus.

▶ Klicken Sie auf OK.

Ordinale Regression: Optionen

Im Dialogfeld "Ordinale Regression: Optionen" können Sie die im iterativen Schätzprozeß verwendeten Parameter anpassen, ein Konfidenzniveau für die Parameterschätzer bestimmen und eine Link-Funktion auswählen.

Abbildung 27-2
Dialogfeld "Ordinale Regression: Optionen"

Iterationen. Sie können den Iterationsprozeß anpassen.

- **Maximale Anzahl der Iterationen.** Geben Sie eine nichtnegative ganze Zahl an. Beim Wert 0 gibt die Prozedur die anfänglichen Schätzwerte zurück.
- **Maximalzahl für Schritt-Halbierung.** Geben Sie eine positive ganze Zahl ein.
- **Log-Likelihood-Konvergenz.** Der Prozeß wird beendet, wenn die absolute oder relative Änderung der Log-Likelihood kleiner als dieser Wert ist. Bei einem Wert von 0 wird dieses Kriterium nicht verwendet.
- **Parameter-Konvergenz.** Der Prozeß wird beendet, wenn die absolute oder relative Änderung in jedem der Parameterschätzer kleiner als dieser Wert ist. Bei einem Wert von 0 wird dieses Kriterium nicht verwendet.

Konfidenzintervall. Geben Sie einen Wert größer oder gleich 0 und kleiner als 100 ein.

Delta. Der Wert, der zu Zellen mit einer Häufigkeit von 0 addiert wird. Geben Sie eine nicht-negative Zahl kleiner als 1 an.

Toleranz für Prüfung auf Singularität. Wird zum Prüfen auf stark abhängige Einflußvariablen verwendet. Wählen Sie einen Wert aus der Liste der Optionen aus.

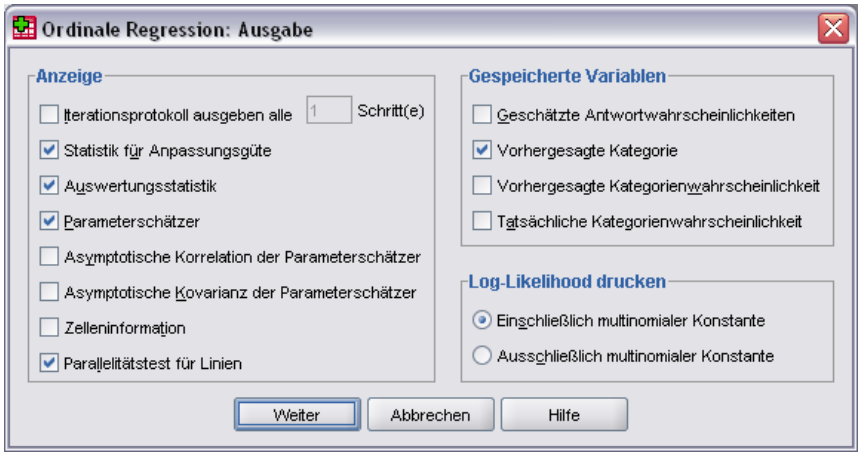
Link-Funktion. Die Link-Funktion ist eine Transformation der kumulativen Wahrscheinlichkeiten, die eine Schätzung des Modells ermöglicht. Es stehen fünf Link-Funktionen zur Verfügung, die in der folgenden Tabelle zusammengefasst sind.

Funktion	Form	Typische Anwendung
Logit	$\log(\xi / (1-\xi))$	Gleichmäßig verteilte Kategorien
Log-Log komplementär	$\log(-\log(1-\xi))$	Höhere Kategorien wahrscheinlicher
Log-Log negativ	$-\log(-\log(\xi))$	Niedrigere Kategorien wahrscheinlicher
Probit	$\Phi^{-1}(\xi)$	Latente Variable ist normalverteilt
Cauchit (Inverse von Cauchy)	$\tan(\pi(\xi-0,5))$	Latente Variable weist viele Extremwerte auf

Ordinale Regression: Ausgabe

Im Dialogfeld "Ordinale Regression: Ausgabe" können Sie festlegen, welche Tabellen im Viewer angezeigt werden und ob Variablen in der Arbeitsdatei gespeichert werden.

Abbildung 27-3
Dialogfeld "Ordinale Regression: Ausgabe"



Anzeigen. Es werden die folgenden Tabellen erstellt:

- **Iterationsprotokoll ausgeben.** Die Log-Likelihood und die Parameterschätzer werden mit der hier angegebenen Häufigkeit ausgegeben. Die erste und letzte Iteration wird immer ausgegeben.
- **Statistik für Anpassungsgüte.** Gibt die Chi-Quadrat-Statistik nach Pearson und die Likelihood-Quotienten-Chi-Quadrat-Statistik aus. Diese werden anhand der in der Variablenliste angegebenen Klassifikation berechnet.
- **Auswertungsstatistik.** R^2-Statistik nach Cox und Snell, Nagelkerke und McFadden.
- **Parameterschätzer.** Parameterschätzer, Standardfehler und Konfidenzintervalle.
- **Asymptotische Korrelation der Parameterschätzer.** Matrix der Parameterschätzer-Korrelationen.
- **Asymptotische Kovarianz der Parameterschätzer.** Matrix der Parameterschätzer-Kovarianzen.
- **Zelleninformationen.** Beobachtete und erwartete Häufigkeiten und kumulative Häufigkeiten, Pearson-Residuen für Häufigkeiten und kumulative Häufigkeiten, beobachtete und erwartete Wahrscheinlichkeiten sowie beobachtete und erwartete kumulative Wahrscheinlichkeiten jeder Antwortkategorie nach Kovariaten-Struktur. Bedenken Sie, daß diese Option bei Modellen mit vielen Kovariaten-Strukturen (beispielsweise bei Modellen mit stetigen Kovariaten) zu einer sehr umfassenden, unübersichtlichen Tabelle führen kann.
- **Parallelitätstest für Linien.** Test der Hypothese, daß die Kategorieparameter über alle Niveaus der abhängigen Variablen gleich sind. Dies ist nur bei reinen Kategoriemodellen verfügbar.

Gespeicherte Variablen. Es werden die folgenden Variablen in der Arbeitsdatei gespeichert:

- **Geschätzte Antwortwahrscheinlichkeiten.** Aus dem Modell geschätzte Wahrscheinlichkeiten, daß eine Faktor-/Kovariaten-Struktur in eine Antwortkategorie klassifiziert wird. Es gibt so viele Wahrscheinlichkeiten wie die Anzahl der Antwortkategorien.

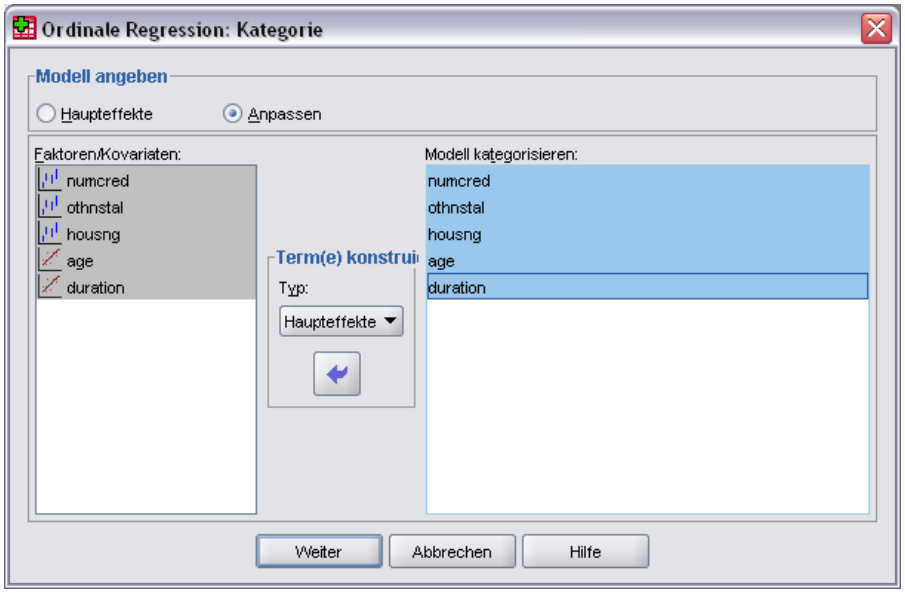
- **Vorhergesagte Kategorie.** Die Antwortkategorie mit der größten geschätzten Wahrscheinlichkeit für eine Faktor-/Kovariaten-Struktur.
- **Vorhergesagte Kategorienwahrscheinlichkeit.** Geschätzte Wahrscheinlichkeit, daß eine Faktor-/Kovariaten-Struktur in die vorhergesagte Kategorie klassifiziert wird. Diese Wahrscheinlichkeit entspricht außerdem der größten geschätzten Wahrscheinlichkeit der Faktor-/Kovariaten-Struktur.
- **Tatsächliche Kategorienwahrscheinlichkeit.** Geschätzte Wahrscheinlichkeit, daß eine Faktor-/Kovariaten-Struktur in die tatsächliche Kategorie klassifiziert wird.

Log-Likelihood drucken. Hiermit wird die Ausgabe der Log-Likelihood festgelegt. Mit Einschließlich multinomialer Konstante wird der vollständige Wert der Likelihood ausgegeben. Wenn Sie die Ergebnisse mit anderen Produkten vergleichen möchten, bei denen keine Konstante vorhanden ist, können Sie diese ausschließen.

Ordinale Regression: Kategorie

Im Dialogfeld "Ordinale Regression: Kategorie" können Sie das Modell für die Analyse kategorisieren.

Abbildung 27-4
Dialogfeld "Ordinale Regression: Kategorie"



Modell bestimmen. Ein Modell mit Haupteffekten enthält die Haupteffekte der Faktoren und Kovariaten, aber keine Wechselwirkungseffekte. Sie können ein benutzerdefiniertes Modell erstellen, um Teilgruppen von Wechselwirkungen zwischen Faktoren oder Kovariaten zu bestimmen.

Faktoren/Kovariaten. Die Faktoren und Kovariaten werden aufgelistet.

Modell kategorisieren. Das Modell ist abhängig von den gewählten Haupt- und Wechselwirkungseffekten.

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-fach. Hiermit werden alle möglichen 2-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-fach. Hiermit werden alle möglichen 3-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

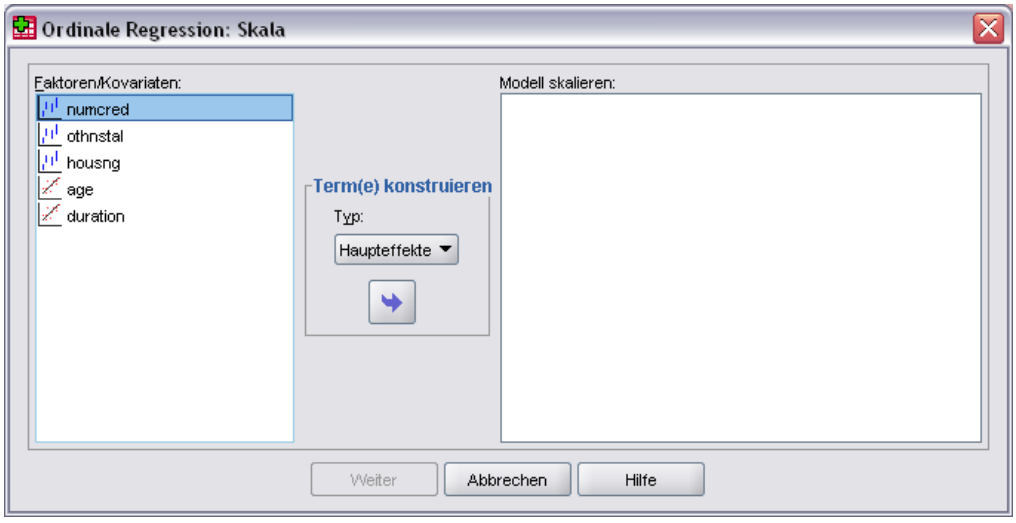
Alle 4-fach. Hiermit werden alle möglichen 4-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-fach. Hiermit werden alle möglichen 5-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Ordinale Regression: Skala

Im Dialogfeld "Ordinale Regression: Skala" können Sie das Modell für die Analyse skalieren.

Abbildung 27-5
Dialogfeld "Ordinale Regression: Skala"



Faktoren/Kovariaten. Die Faktoren und Kovariaten werden aufgelistet.

Modell skalieren. Das Modell ist abhängig von den gewählten Haupt- und Wechselwirkungseffekten.

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-fach. Hiermit werden alle möglichen 2-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 3-fach. Hiermit werden alle möglichen 3-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-fach. Hiermit werden alle möglichen 4-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-fach. Hiermit werden alle möglichen 5-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Zusätzliche Funktionen beim Befehl PLUM

Sie können die ordinale Regression an Ihre Bedürfnisse anpassen, wenn Sie ihre Auswahl in ein Syntax-Fenster einfügen und die resultierende Befehlssyntax für den Befehl `PLUM` bearbeiten. Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Angepasste Hypothesentests können durch Festlegen von Nullhypothesen als lineare Parameterkombinationen erstellt werden.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
28

Kurvenanpassung

Mit der Prozedur “Kurvenanpassung” werden Regressionsstatistiken zur Kurvenanpassung und zugehörige Diagramme für 11 verschiedene Regressionsmodelle zur Kurvenanpassung erstellt. Für jede abhängige Variable wird ein separates Modell erstellt. Außerdem können Sie vorhergesagte Werte, Residuen und Vorhersageintervalle als neue Variablen speichern.

Beispiel. Ein Internet-Dienstanbieter verfolgt den Prozentsatz des mit Viren infizierten E-Mail-Verkehrs über die Netzwerke im Lauf der Zeit. Ein Streudiagramm zeigt, dass eine nichtlineare Beziehung vorliegt. Sie können ein quadratisches oder kubisches Modell an die Daten anpassen und die Gültigkeit der Annahmen sowie die Güte der Anpassung des Modells prüfen.

Statistiken. Für jedes Modell: Regressionskoeffizienten, multiples R, R^2, korrigiertes R^2, Standardfehler des Schätzers, Tabelle für die Varianzanalyse, vorhergesagte Werte, Residuen und Vorhersageintervalle. Modelle: linear, logarithmisch, invers, quadratisch, kubisch, Potenz, zusammengesetzt, S-Kurve, logistisch, Wachstum und exponentiell.

Daten. Die abhängigen und die unabhängigen Variablen müssen quantitativ sein. Wenn Sie aus der Arbeitsdatei Zeit als unabhängige Variable ausgewählt haben (statt eine Variable auszuwählen), erzeugt die Prozedur “Kurvenanpassung” eine Zeitvariable mit gleichen Zeitabständen zwischen den Fällen. Wenn Zeit ausgewählt wurde, sollte die abhängige Variable eine Zeitreihenmessung sein. Zur Zeitreihenanalyse ist eine Datendateistruktur erforderlich, in der jeder Fall (jede Zeile) einen Satz von Beobachtungen zu unterschiedlichen Zeiten bei gleichen Zeitabständen zwischen den Fällen darstellt.

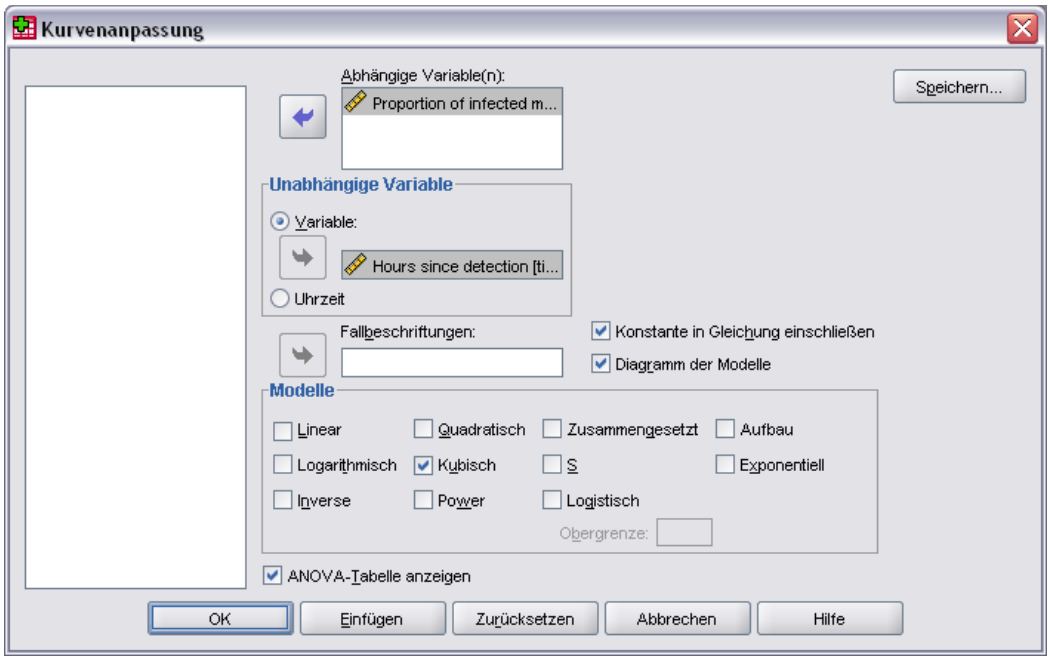
Annahmen. Stellen Sie Ihre Daten grafisch dar, um den Zusammenhang zwischen den unabhängigen und den abhängigen Variablen (linear, exponentiell usw.) erkennen zu können. Die Residuen eines guten Modells müssen willkürlich und normalverteilt sein. Bei einem linearen Modell müssen folgende Annahmen erfüllt werden: Für jeden Wert der unabhängigen Variablen muss die abhängige Variable normalverteilt vorliegen. Die Varianz der Verteilung der abhängigen Variablen muss für alle Werte der unabhängigen Variablen konstant sein. Die abhängige Variable und die unabhängige Variable müssen linear zusammenhängen, und alle Beobachtungen müssen unabhängig sein.

So führen Sie eine Kurvenanpassung durch:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Regression
  Kurvenanpassung…

Abbildung 28-1
Dialogfeld "Kurvenanpassung"



- Wählen Sie eine oder mehrere abhängige Variablen aus. Für jede abhängige Variable wird ein separates Modell erstellt.
- Wählen Sie eine unabhängige Variable aus (wählen Sie entweder eine Variable aus der Arbeitsdatei oder wählen Sie Zeit aus).
- Die folgenden Optionen sind verfügbar:
  - Eine Variable zum Beschriften der Fälle in Streudiagrammen auswählen. Sie können für jeden Punkt im Streudiagramm das Symbol zum Identifizieren von Punkten verwenden, um den Wert der Variablen für die "Fallbeschriftung" anzeigen zu lassen.
  - Klicken Sie auf Speichern, um vorhergesagte Werte, Residuen und Vorhersageintervalle als neue Variablen zu speichern.

Außerdem sind folgende Optionen verfügbar:

- **Konstante in Gleichung einschließen.** Mit dieser Option wird ein konstanter Term in der Regressionsgleichung geschätzt. In der Standardeinstellung ist die Konstante eingeschlossen.
- **Diagramm der Modelle.** Mit dieser Option werden für alle ausgewählten Modelle die Werte der abhängigen Variablen über der unabhängigen Variablen grafisch dargestellt. Für jede abhängige Variable wird ein eigenes Diagramm erzeugt.
- **ANOVA-Tabelle anzeigen.** Mit dieser Option wird für jedes ausgewählte Modell eine Zusammenfassung für die Varianzanalyse angezeigt.

Modelle für die Kurvenanpassung

Sie können ein oder mehrere Regressionsmodelle für die Kurvenanpassung auswählen. Stellen Sie Ihre Daten grafisch dar, um zu ermitteln, welches Modell Sie verwenden sollten. Wenn Ihre Variablen in einem linearen Zusammenhang zu stehen scheinen, verwenden Sie ein einfaches lineares Regressionsmodell. Wenn Ihre Variablen in keinem linearen Zusammenhang stehen, transformieren Sie diese. Wenn eine Transformation keine Abhilfe schafft, benötigen Sie möglicherweise ein komplizierteres Modell. Betrachten Sie ein Streudiagramm Ihrer Daten. Wenn das Diagramm einer Ihnen bekannten mathematischen Funktion ähnelt, passen Sie Ihre Daten an diesen Modelltyp an. Wenn Ihre Daten zum Beispiel einer Exponentialfunktion ähneln, verwenden Sie ein exponentielles Modell.

Linear. Ein Modell mit der Gleichung Y = b0 + (b1 * t). Die Werte der Zeitreihe werden als lineare Funktion der Zeit aufgefasst.

Logarithmisch. Ein Modell mit der Gleichung Y = b0 + (b1 * ln(t)).

Inverse. Ein Modell mit der Gleichung Y = b0 + (b1 / t).

Quadratisch. Ein Modell mit folgender Gleichung: Y = b0 + (b1 * t) + (b2 * t**2). Das quadratische Modell kann zum Modellieren von Zeitreihen verwendet werden, die "abheben" oder gedämpft verlaufen.

Kubisch. Ein Modell mit folgender Gleichung:Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3).

Power. Ein Modell mit folgender Gleichung: Y = b0 * (t**b1) oder ln(Y) = ln(b0) + (b1 * ln(t)).

Zusammengesetzt. Dieses Modell basiert auf folgender Gleichung: Y = b0 * (b1**t) oder ln(Y) = ln(b0) + (ln(b1) * t).

S-Kurve. Ein Modell, dessen Gleichung lautet: Y = e**(b0 + (b1/t)) oder ln(Y) = b0 + (b1/t).

Logistisch. Die Gleichung für dieses Modell lautet Y = 1 / (1/u + (b0 * (b1**t))) oder ln(1/y-1/u) = ln (b0) + (ln(b1) * t), wobei u die obere Schranke ist. Nach der Auswahl von "Logistisch" muss der Wert der oberen Schranke angegeben werden, der in der Regressionsgleichung verwendet werden soll. Der Wert muss eine positive Zahl sein, die größer ist als der größte Wert der abhängigen Variablen.

Aufbau. Ein Modell, dessen Gleichung lautet: Y = e**(b0 + (b1 * t)) oder ln (Y) = b0 + (b1 * t).

Exponentiell. Ein Modell mit folgender Gleichung: Y = b0 * (e**(b1 * t)) oder ln (Y) = ln (b0) + (b1 * t).

Kurvenanpassung: Speichern

Abbildung 28-2
Dialogfeld "Kurvenanpassung: Speichern"

Variablen speichern. Für jedes ausgewählte Modell können Sie vorhergesagte Werte, Residuen (beobachteter Wert der abhängigen Variablen minus vorhergesagter Wert des Modells) und Vorhersageintervalle (Ober- und Untergrenzen) speichern. Die neuen Variablennamen werden mit den beschreibenden Labels in einer Tabelle im Ausgabefenster angezeigt.

Fälle vorhersagen. Wenn Sie in der Arbeitsdatei statt einer Variablen Zeit als unabhängige Variable ausgewählt haben, können Sie nach dem Ende der Zeitreihe eine Vorhersageperiode angeben. Sie können eine der folgenden Möglichkeiten wählen:

- **Von der Schätzperiode bis zum letzten Fall vorhersagen.** Hiermit werden auf der Grundlage der Fälle in der Schätzperiode Werte für alle Fälle in der Datei vorhergesagt. Die unten im Dialogfeld angezeigte Schätzperiode wird im Menü "Daten", Option "Fälle auswählen", Dialogfeld "Fälle auswählen:Bereich" festgelegt. Wenn keine Schätzperiode definiert wurde, werden alle Fälle zum Schätzen der Werte verwendet.
- **Vorhersagen bis.** Hiermit werden auf der Grundlage der Fälle in der Schätzperiode Werte bis zum angegebenen Datum, zur angegebenen Uhrzeit oder zur angegebenen Beobachtungsnummer vorhergesagt. Mit dieser Funktion können Werte nach dem letzten Fall in der Zeitreihe vorhergesagt werden. Die gegenwärtig definierten Datumsvariablen bestimmen, welche Textfelder zur Verfügung stehen, um das Ende der Vorhersageperiode anzugeben. Wenn keine Datumsvariablen definiert sind, können Sie die letzte Beobachtungs- bzw. Fallnummer angeben.

Datumsvariablen erstellen Sie im Menü "Daten" mit der Option "Datum definieren".

Kapitel
29

Regression mit partiellen kleinsten Quadraten

Die Prozedur “Regression mit partiellen kleinsten Quadraten” schätzt Regressionsmodelle mit partiellen kleinsten Quadraten (Partial Least Squares, PLS auch als “Projektion auf latente Struktur” (Projection to Latent Structure) bezeichnet). PLS ist ein Vorhersageverfahren, das eine Alternative zum Regressionsmodell der gewöhnlichen kleinsten Quadrate (Ordinary Least Squares, OLS), zur kanonischen Korrelation bzw. zur Modellierung von Strukturgleichungen darstellt und besonders nützlich ist, wenn die Einflussvariablen eine hohe Korrelation aufweisen oder wenn die Anzahl der Einflussvariablen die Anzahl der Fälle übersteigt.

PLS kombiniert Merkmale der Hauptkomponentenanalyse mit Merkmalen der multiplen Regression. Zunächst wird ein Set latenter Faktoren extrahiert, die einen möglichst großen Anteil der Kovarianz zwischen den unabhängigen und den abhängigen Variablen erklären. Anschließend werden in einem Regressionsschritt die Werte der abhängigen Variablen mithilfe der Zerlegung der unabhängigen Variablen vorhergesagt.

Verfügbarkeit. PLS ist ein Erweiterungsbefehl, für den das Python-Erweiterungsmodul auf dem System installiert sein muss, auf dem PLS ausgeführt werden soll. Das PLS-Erweiterungsmodul muss separat installiert werden. Das Installationsprogramm können Sie auf folgender Webseite herunterladen: *http://www.spss.com/devcentral*.

Anmerkung: Das PLS-Erweiterungsmodul ist von Python-Software abhängig. SPSS ist nicht der Inhaber bzw. Lizenzgeber der Python-Software. Alle Python-Benutzer müssen den Bestimmungen der Python-Lizenzvereinbarung zustimmen, die sich auf der Python-Website befindet. SPSS gibt keinerlei Erklärungen über die Qualität des Python-Programms ab. SPSS übernimmt keinerlei Haftung in Zusammenhang mit Ihrer Verwendung des Python-Programms.

Tabellen. Der Anteil der (durch den latenten Faktor) erklärten Varianz, die Gewichtungen latenter Faktoren, die Ladungen latenter Faktoren, die Bedeutung der unabhängigen Variablen in der Projektion (VIP) und die Schätzer für Regressionsparameter (nach abhängiger Variablen) werden jeweils standardmäßig angegeben.

Diagramme. Die Bedeutung der Variablen in der Projektion (Variable Importance in Projection, VIP), Faktor-Scores, Faktorgewichtungen für die ersten drei latenten Faktoren und die Distanz zum Modell werden jeweils über die Registerkarte Optionen erstellt.

Messniveau. Die abhängigen und unabhängigen Variablen (Einflussvariablen) können metrisch, nominal oder ordinal sein. Bei der Prozedur wird davon ausgegangen, dass allen Variablen das richtige Messniveau zugewiesen wurde. Sie können das Messniveau für eine Variable jedoch vorübergehend ändern. Klicken Sie hierzu mit der rechten Maustaste auf die Variable in der Liste der Quellvariablen und wählen Sie das gewünschte Messniveau im Kontextmenü aus. Kategoriale (nominale bzw. ordinale) Variablen werden von der Prozedur gleich behandelt.

Kodierung für kategoriale Variablen. Die Prozedur kodiert vorübergehend für die Dauer des Verfahrens kategoriale abhängige Variablen mithilfe der "Eins-aus-*c*"-Kodierung neu. Wenn es *c* Kategorien für eine Variable gibt, wird die Variable als *c* Vektoren gespeichert. Dabei wird die erste Kategorie als (1,0,...,0) angegeben, die zweite Kategorie als (0,1,0,...,0), ... und die letzte Kategorie als (0,0,...,0,1). Kategoriale abhängige Variablen werden mithilfe von Dummy-Codierung dargestellt, d. h. es wird einfach der Indikator weggelassen, der der Referenzkategorie entspricht.

Häufigkeitsgewichtungen. Gewichtungswerte werden vor der Verwendung auf die nächste ganze Zahl gerundet. Fälle mit fehlenden Gewichten oder Gewichten unter 0,5 werden in der Analyse nicht verwendet.

Fehlende Werte. Benutzer- und systemdefiniert fehlende Werte werden als ungültig behandelt.

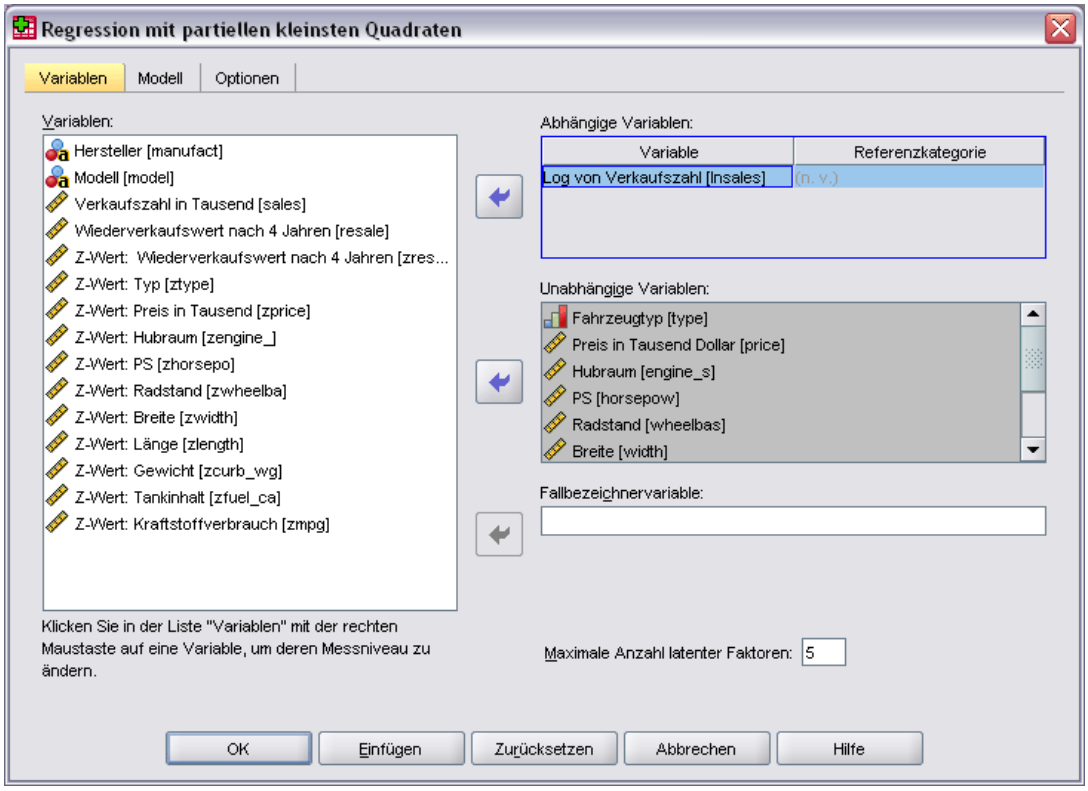
Neuskalierung. Alle Modellvariablen werden zentriert und standardisiert, einschließlich der Indikatorvariablen die für kategoriale Variablen stehen.

So lassen Sie eine Regression mit partiellen kleinsten Quadraten berechnen:

Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Regression
    Partielle kleinste Quadrate...

Abbildung 29-1
Regression mit partiellen kleinsten Quadraten – Registerkarte "Variablen"



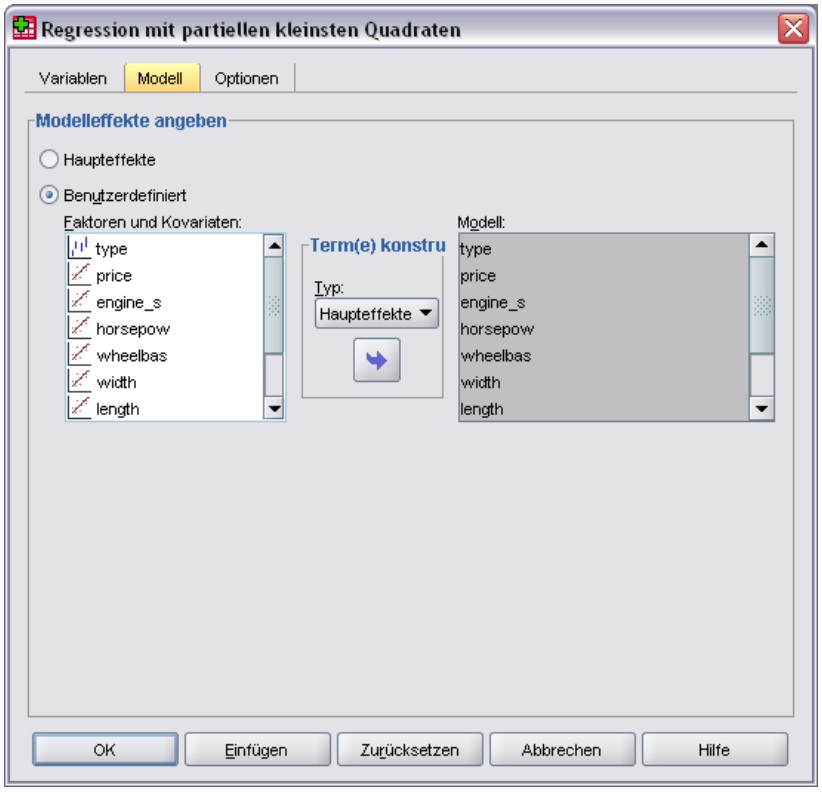
- ▶ Wählen Sie mindestens eine abhängige Variable aus.
- ▶ Wählen Sie mindestens eine unabhängige Variable aus.

Die folgenden Optionen sind verfügbar:

- Angabe einer Referenzkategorie für kategoriale (nominale bzw. ordinale) abhängige Variablen.
- Angabe einer Variablen, die als eindeutige Kennung für die fallweise Ausgabe und für die gespeicherten Daten-Sets verwendet werden soll.
- Angabe einer Obergrenze für die Anzahl der zu extrahierenden latenten Faktoren.

Modell

Abbildung 29-2
Regression mit partiellen kleinsten Quadraten – Registerkarte "Modell"



Modell-effekte angeben. Ein Modell mit Haupteffekten enthält die Haupteffekte aller Faktoren und Kovariaten. Wählen Sie Benutzerdefiniert, um Interaktionen anzugeben. Sie müssen alle in das Modell zu übernehmenden Terme angeben.

Faktoren und Kovariaten. Die Faktoren und Kovariaten werden aufgelistet.

Modell. Das Modell ist von der Art Ihrer Daten abhängig. Nach der Auswahl von Anpassen können Sie die Haupteffekte und Wechselwirkungen auswählen, die für Ihre Analyse von Interesse sind.

Terme konstruieren

Für die ausgewählten Faktoren und Kovariaten:

Wechselwirkung. Hiermit wird der Wechselwirkungsterm mit der höchsten Ordnung von allen ausgewählten Variablen erzeugt. Dies ist die Standardeinstellung.

Haupteffekte. Legt einen Haupteffekt-Term für jede ausgewählte Variable an.

Alle 2-fach. Hiermit werden alle möglichen 2-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

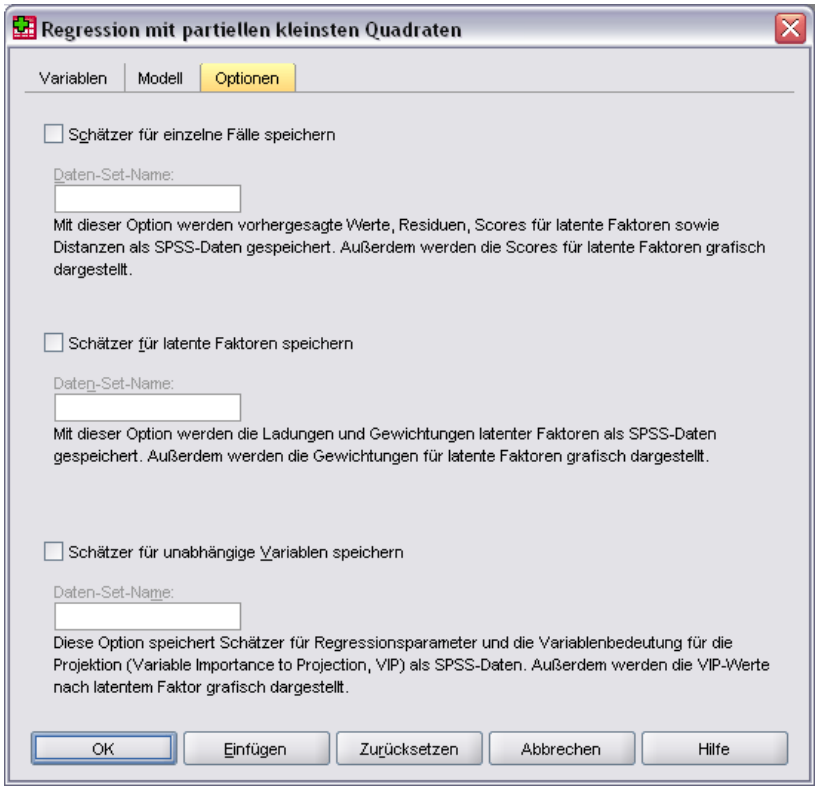
Alle 3-fach. Hiermit werden alle möglichen 3-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 4-fach. Hiermit werden alle möglichen 4-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Alle 5-fach. Hiermit werden alle möglichen 5-fach-Wechselwirkungen der ausgewählten Variablen erzeugt.

Optionen

Abbildung 29-3
Regression mit partiellen kleinsten Quadraten – Registerkarte "Optionen"



Auf der Registerkarte "Optionen" kann der Benutzer Modellschätzer für einzelne Fälle, latente Faktoren und Einflussvariablen speichern und grafisch darstellen lassen.

Geben Sie für jeden Datentyp den Namen eines SPSS-Daten-Sets an. Die Namen der Daten-Sets müssen eindeutig sein. Wenn Sie den Namen eines bestehenden Daten-Sets angeben, werden dessen Inhalte ersetzt; ansonsten wird ein neues Daten-Set erstellt.

- **Schätzer für einzelne Fälle speichern.** Speichert die folgenden fallweisen Modellschätzer: vorhergesagte Werte, Residuen, Distanz zum Modell mit latenten Faktoren und Scores für latente Faktoren. Außerdem werden die Scores für latente Faktoren grafisch dargestellt.
- **Schätzer für latente Faktoren speichern.** Speichert die Ladungen und Gewichtungen latenter Faktoren. Außerdem werden die Gewichtungen für latente Faktoren grafisch dargestellt.
- **Schätzer für unabhängige Variablen speichern.** Speichert Schätzer für Regressionsparameter und die Bedeutung der unabhängigen Variablen in der Projektion (VIP). Außerdem werden die VIP-Werte für die einzelnen latente Faktoren grafisch dargestellt.

Kapitel
30

Diskriminanzanalyse

Die Diskriminanzanalyse dient zur Erstellung eines Vorhersagemodells der Gruppenzugehörigkeit. Das Modell besteht aus einer Diskriminanzfunktion (oder, bei mehr als zwei Gruppen, einem Set von Diskriminanzfunktionen) auf der Grundlage derjenigen linearen Kombinationen der Prädiktorvariablen, welche die beste Diskriminanz zwischen den Gruppen ergeben. Die Funktionen werden aus einer Stichprobe der Fälle erzeugt, bei denen die Gruppenzugehörigkeit bekannt ist. Diese Funktionen können dann auf neue Fälle mit Messungen für die Prädiktorvariablen, aber unbekannter Gruppenzugehörigkeit angewandt werden.

Anmerkung: Die Gruppenvariable kann mehr als zwei Werte besitzen. Die Codes für die Gruppenvariable müssen allerdings ganzzahlige Werte sein, und Sie müssen hierfür die minimalen und maximalen Werte festlegen. Fälle mit Werten außerhalb dieser Grenzen werden von der Analyse ausgeschlossen.

Beispiel. Im Durchschnitt verbrauchen Personen in kühlen Ländern mehr Kalorien pro Tag als Bewohner der Tropen, und ein größerer Anteil der Personen in den kühlen Ländern sind Stadtbewohner. Ein Forscher möchte diese Informationen in einer Funktion zusammenfassen, um zu bestimmen, wie gut eine bestimmte Person diesen beiden Ländergruppen zugeordnet werden kann. Der Forscher nimmt an, dass auch die Bevölkerungsgröße und Wirtschaftsinformationen relevant sein könnten. Mit der Diskriminanzanalyse können Sie die Koeffizienten der linearen Diskriminanzfunktion schätzen, die im Prinzip genauso wie die rechte Seite einer Regressionsgleichung bei mehrfacher Regression aufgebaut ist. Unter Verwendung der Koeffizienten *a*, *b*, *c* und *d* lautet die Funktion also:

D = a * Klima + b * Städtisch + c * Bevölkerung + d * Bruttosozialprodukt der Region je Einwohner.

Wenn diese Variablen für die Unterscheidung zwischen den beiden Klimazonen relevant sind, müssen sich die Werte von *D* für tropische und kühlere Länder unterscheiden. Falls Sie eine schrittweise Methode für die Variablenauswahl verwenden, stellen Sie unter Umständen fest, dass nicht alle vier Variablen in die Funktion aufgenommen werden müssen.

Statistiken. Für jede Variable: Mittelwerte, Standardabweichungen, univariate ANOVA. Für jede Analyse: Box-*M*, Korrelationsmatrix innerhalb der Gruppen, Kovarianzmatrix innerhalb der Gruppen, Kovarianzmatrix der einzelnen Gruppen, gesamte Kovarianzmatrix. Für jede kanonische Diskriminanzfunktion: Eigenwert, Prozentwert der Varianz, kanonische Korrelation, Wilks-Lambda, Chi-Quadrat. Für jeden Schritt: a-priori-Wahrscheinlichkeit, Funktionskoeffizienten nach Fisher, nicht standardisierte Funktionskoeffizienten, Wilks-Lambda für jede kanonische Funktion.

Daten. Die Gruppenvariable muss über eine begrenzte Anzahl unterschiedener Kategorien verfügen, die als ganzzahlige Werte kodiert werden. Unabhängige nominale Variablen müssen in Dummy- oder Kontrastvariablen umkodiert werden.

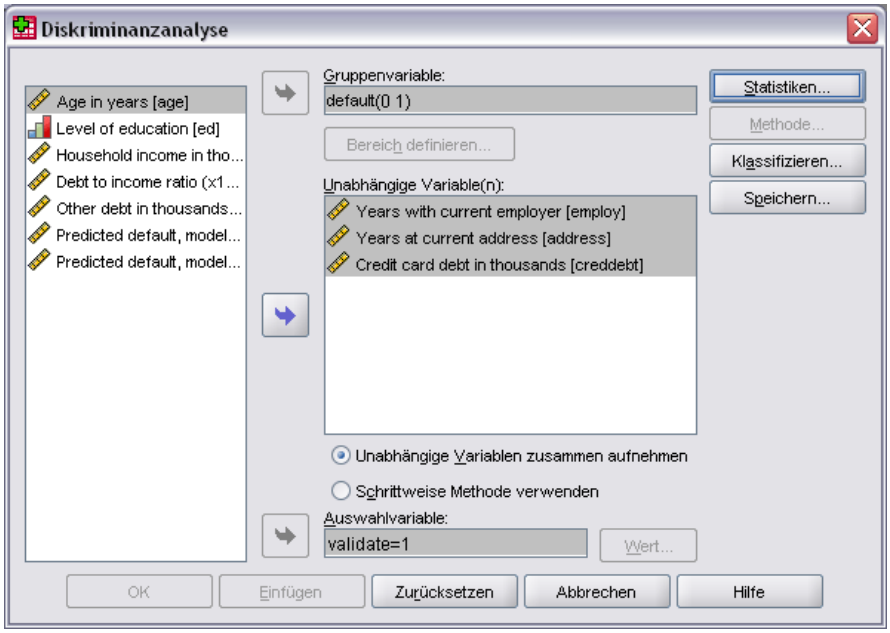
Annahmen. Die Fälle müssen unabhängig sein. Einflußvariablen müssen in multivariater Normalverteilung vorliegen, und die Varianz-Kovarianz-Matrizen innerhalb der Gruppen müssen zwischen den Gruppen gleich groß sein. Die Gruppenzugehörigkeit muss sich wechselseitig ausschließen (das heißt, kein Fall gehört zu mehr als einer Gruppe) und umfassend sein (das heißt, alle Fälle gehören zu einer Gruppe). Diese Prozedur ist am effektivsten, wenn die Gruppenzugehörigkeit eine rein kategoriale Variable ist. Wenn die Gruppenzugehörigkeit hingegen auf den Werten einer stetigen Variablen basiert (zum Beispiel bei einem Vergleich von IQ-Werten), sollten Sie die lineare Regression in Betracht ziehen, um von den reichhaltigeren Informationen zu profitieren, die in der stetigen Variablen selbst enthalten sind.

So lassen Sie eine Diskriminanzanalyse berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Klassifizieren
    Diskriminanzanalyse...

Abbildung 30-1
Dialogfeld "Diskriminanzanalyse"



▶ Wählen Sie eine Gruppenvariable mit ganzzahligen Werten aus und klicken Sie auf Bereich definieren, um die gewünschten Kategorien festzulegen.

▶ Wählen Sie die unabhängigen Variablen oder Einflußvariablen aus. (Wenn die Gruppenvariable nichtganzzahlig ist, können Sie eine Variable mit dieser Eigenschaft im Menü "Transformieren" mit dem Befehl "Automatisch umkodieren" erstellen.)

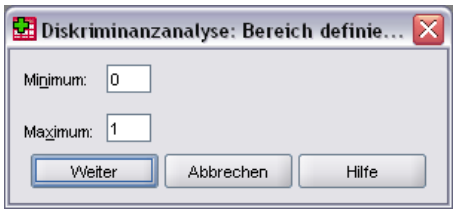
▶ Wählen Sie die gewünschte Methode für die Eingabe der unabhängigen Variablen aus.

- **Unabhängige Variablen zusammen aufnehmen.** Nimmt alle unabhängigen Variablen, welche die Toleranzkriterien erfülllen, gleichzeitig auf.
- **Schrittweise Methode verwenden.** Verwendet ein schrittweises Verfahren zur Steuerung von Variablenaufnahme und Variablenausschluss.

▶ Wahlweise können Sie die Fälle auch mithilfe einer Auswahlvariablen auswählen.

Diskriminanzanalyse: Bereich definieren

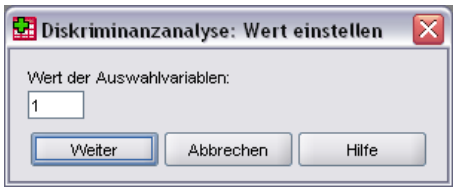
Abbildung 30-2
Dialogfeld "Diskriminanzanalyse: Bereich definieren"



Geben Sie den kleinsten (Minimum) und den größten (Maximum) Wert der Gruppenvariablen für die Analyse an. Fälle mit Werten außerhalb dieses Bereichs werden in der Diskriminanzanalyse nicht verwendet, aber ausgehend von den Ergebnissen der Analyse in eine der vorhandenen Gruppen eingeordnet. Die Minimum- und Maximumwerte müssen ganzzahlig sein.

Diskriminanzanalyse: Fälle auswählen

Abbildung 30-3
Dialogfeld "Diskriminanzanalyse: Wert einstellen"



So wählen Sie die Fälle für die Analyse aus:

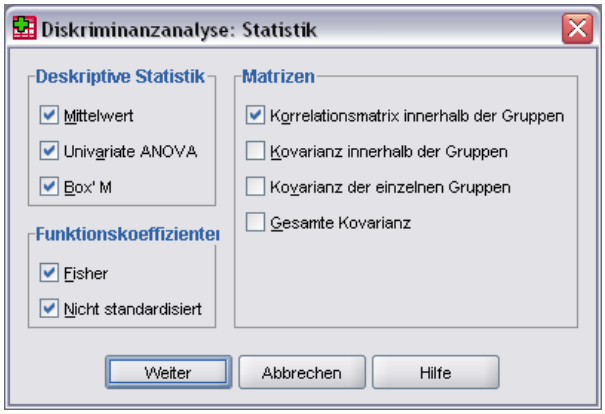
▶ Wählen Sie im Dialogfeld "Diskriminanzanalyse" eine Auswahlvariable aus.

▶ Klicken Sie auf Wert, um eine ganze Zahl als Auswahlvariable einzugeben.

Bei der Ableitung der Diskriminanzfunktionen werden nur die Fälle verwendet, deren Auswahlvariablen den angegebenen Wert aufweisen. Statistiken und Klassifikationsergebnisse werden sowohl für die ausgewählten als auch für die nicht ausgewählten Fälle erzeugt. Mit diesem Prozess liegt ein Mechanismus vor, mit dem neue Fälle anhand von bereits vorhandenen Daten klassifiziert werden können oder mit dem Sie Ihre Daten in Teilmengen von Lern- und Testfällen einteilen können, um so eine Gültigkeitsprüfung des erzeugten Modells durchzuführen.

Diskriminanzanalyse: Statistik

Abbildung 30-4
Dialogfeld "Diskriminanzanalyse: Statistik"



Deskriptive Statistiken. Verfügbare Optionen sind Mittelwerte (einschließlich Standardabweichungen), univariate ANOVA und Box' *M*-Test.

- **Mittelwerte.** Zeigt Gesamt- und Gruppenmittelwerte sowie Standardabweichungen für die unabhängigen Variablen an.
- **Univariate ANOVA.** Führt für jede unabhängige Variable eine einfaktorielle Varianzanalyse durch, d. h. einen Test auf Gleichheit der Gruppenmittelwerte.
- **Box-M.** Ein Test auf Gleichheit der Kovarianzmatrizen der Gruppen. Bei hinreichend großen Stichproben bedeutet ein nichtsignifikanter p-Wert, dass die Anhaltspunkte für unterschiedliche Matrizen nicht ausreichend sind. Der Test ist empfindlich gegenüber Abweichungen von der multivariaten Normalverteilung.

Funktionskoeffizienten. Verfügbare Optionen sind Klassifikationskoeffizienten nach Fisher und nicht standardisierte Koeffizienten.

- **Fisher.** Zeigt die Koeffizienten der Klassifizierungsfunktion nach Fisher an, die direkt für die Klassifizierung verwendet werden können. Es wird ein Set von Koeffizienten für jede Gruppe ermittelt. Ein Fall wird der Gruppe zugewiesen, für den er den größten Diskriminanzwert aufweist.
- **Nichtstandardisiert.** Zeigt die nichtstandardisierten Koeffizienten der Diskriminanzfunktion an.

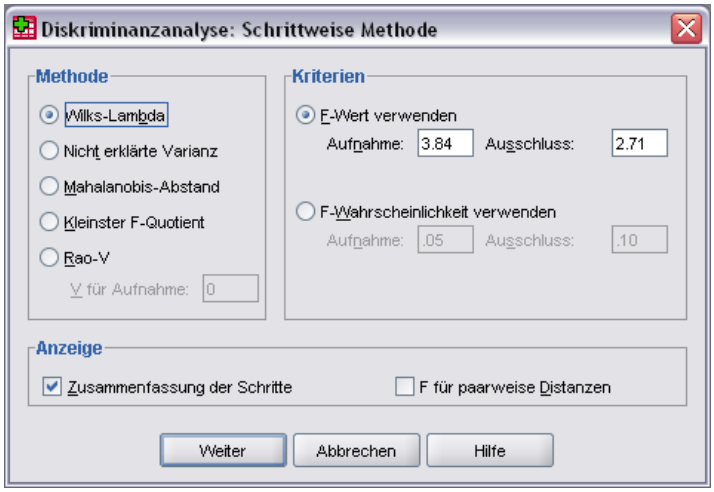
Matrizen. Als Koeffizientenmatrizen für unabhängige Variablen stehen die Korrelationsmatrix innerhalb der Gruppen, die Kovarianzmatrix innerhalb der Gruppen, die gruppenspezifische Kovarianzmatrix und die Kovarianzmatrix für alle Fälle zur Verfügung.

- **Korrelationsmatrix innerhalb der Gruppen.** Zeigt eine gemeinsame Korrelationsmatrix innerhalb der Gruppen an, die als Mittel der separaten Kovarianzmatrizen für alle Gruppen vor der Berechnung der Korrelationen bestimmt wird.
- **Kovarianzmatrix innerhalb der Gruppen.** Zeigt eine gemeinsame Kovarianzmatrix innerhalb der Gruppen an, die sich von der Gesamt-Kovarianzmatrix unterscheiden kann. Die Matrix wird als Mittel der einzelnen Kovarianzmatrizen für alle Gruppen berechnet.

- **Gruppenspezifische Kovarianzmatrix.** Zeigt separate Kovarianzmatrizen für jede Gruppe an.
- **Kovarianzmatrix für alle Fälle.** Zeigt die Kovarianzmatrix für alle Fälle an, so als wären sie aus einer einzigen Stichprobe.

Diskriminanzanalyse: Schrittweise Methode

Abbildung 30-5
Dialogfeld "Diskriminanzanalyse: Schrittweise Methode"



Methode. Wählen Sie die Statistiken aus, die für die Aufnahme oder den Ausschluss neuer Variablen dienen sollen. Die Optionen Wilks-Lambda, nicht erklärte Varianz, Mahalanobis-Abstand, kleinster *F*-Quotient und Rao-*V* stehen zur Verfügung. Mit Rao-*V* können Sie den Mindestanstieg von *V* für eine einzugebende Variable angeben.

- **Wilks-Lambda.** Eine Auswahlmethode für Variablen bei der schrittweisen Diskriminanzanalyse. Die Aufnahme von Variablen in die Gleichung erfolgt anhand der jeweiligen Verringerung von Wilks-Lambda. Bei jedem Schritt wird diejenige Variable aufgenommen, die den Gesamtwert von Wilks-Lambda am meisten vermindert.
- **Nicht erklärte Varianz.** Bei jedem Schritt wird die Variable aufgenommen, welche die Summe der nicht erklärten Variation zwischen den Gruppen minimiert.
- **Mahalanobis-Abstand.** Dieses Maß gibt an, wie weit die Werte der unabhängigen Variablen eines Falles vom Mittelwert aller Fälle abweichen. Ein großer Mahalanobis-Abstand charakterisiert einen Fall, der bei einer oder mehreren unabhängigen Variablen Extremwerte besitzt.
- **Kleinster F-Quotient.** Eine Methode für die Variablenauswahl in einer schrittweisen Analyse. Sie beruht auf der Maximierung eines F-Quotienten, der aus dem Mahalanobis-Abstand zwischen den Gruppen errechnet wird.
- **Rao-V.** Ein Maß für die Unterschiede zwischen Gruppenmittelwerten. Auch Lawley-Hotelling-Spur genannt. Bei jedem Schritt wird die Variable aufgenommen, die den Anstieg des Rao-V maximiert. Wenn Sie diese Option ausgewählt haben, geben Sie den Minimalwert ein, den eine Variable für die Aufnahme in die Analyse aufweisen muss.

Kriterien. Verfügbar sind F-Wert verwenden und F-Wahrscheinlichkeit verwenden. Geben Sie Werte für die Aufnahme und den Ausschluss der Variablen an.

- **F-Wert verwenden.** Eine Variable wird in ein Modell aufgenommen, wenn ihr F-Wert größer ist als der Aufnahmewert. Sie wird ausgeschlossen, wenn der F-Wert kleiner ist als der Ausschlusswert. Der Aufnahmewert muss größer sein als der Ausschlusswert und beide Werte müssen positiv sein. Um mehr Variablen in das Modell aufzunehmen, senken Sie den Aufnahmewert. Um mehr Variablen aus dem Modell auszuschließen, erhöhen Sie den Ausschlusswert.
- **Wahrscheinlichkeit von F verwenden.** Eine Variable wird in das Modell aufgenommen, wenn das Signifikanzniveau ihres F-Werts kleiner ist als der Aufnahmewert. Sie wird ausgeschlossen, wenn das Signifikanzniveau größer ist als der Ausschlusswert. Der Aufnahmewert muss kleiner sein als der Ausschlusswert und beide Werte müssen positiv sein. Um mehr Variablen in das Modell aufzunehmen, erhöhen Sie den Aufnahmewert. Um mehr Variablen aus dem Modell auszuschließen, senken Sie den Ausschlusswert.

Anzeigen. Mit Zusammenfassung der Schritte können Sie nach jedem Schritt die Statistiken für alle Variablen anzeigen lassen. Bei Auswahl von F für paarweise Distanzen wird für jedes Gruppenpaar eine Matrix des paarweisen *F*-Quotienten angezeigt.

Diskriminanzanalyse: Klassifizieren

Abbildung 30-6
Diskriminanzanalyse – Dialogfeld "Klassifizieren"

A-priori-Wahrscheinlichkeiten. Diese Option bestimmt, ob die Klassifikationskoeffizienten für A-priori-Kenntnis über Gruppenzugehörigkeiten angepasst werden.

- **Alle Gruppen gleich.** Es wird von gleichen A-priori-Wahrscheinlichkeiten für alle Gruppen ausgegangen; dies hat keine Auswirkungen auf die Koeffizienten.
- **Von Gruppengrößen berechnen.** Die beobachteten Gruppengrößen in Ihrem Beispiel bestimmen die A-priori-Wahrscheinlichkeiten der Gruppenzugehörigkeit. Falls beispielsweise 50 % der in der Analyse aufgenommenen Beobachtungen in die erste, 25 % in die zweite und 25 % in die dritte Gruppe fallen, werden die Klassifikationskoeffizienten angepasst, um die

Wahrscheinlichkeit der Zugehörigkeit zur ersten Gruppe in Bezug auf die anderen beiden Gruppen zu erhöhen.

Anzeigen. Die verfügbaren Anzeigeoptionen lauten: "Fallweise Ergebnisse", "Zusammenfassende Tabelle" und "Klassifikation mit Fallauslassung".

- **Fallweise Ergebnisse.** Für jeden Fall werden Codes für die tatsächliche Gruppe, die vorhergesagte Gruppe, A-posteriori-Wahrscheinlichkeiten und Diskriminanzwerte angezeigt.
- **Zusammenfassende Tabelle.** Die Anzahl der Fälle, die auf Grundlage der Diskriminanzanalyse jeder der Gruppen richtig oder falsch zugeordnet werden. Zuweilen auch als Klassifikationsmatrix bezeichnet.
- **Klassifikation mit Fallauslassung.** Jeder Fall der Analyse wird durch Funktionen aus allen anderen Fällen unter Auslassung dieses Falls klassifiziert. Diese Klassifikation wird auch als "U-Methode" bezeichnet.

Fehlende Werte durch Mittelwert ersetzen. Wenn Sie diese Option wählen, werden fehlende Werte durch den Mittelwert der jeweiligen unabhängigen Variablen ersetzt, allerdings nur während der Klassifikation der Gruppen.

Kovarianzmatrix verwenden. Sie können wählen, ob zur Klassifikation der Fälle die Kovarianzmatrix innerhalb der Gruppen oder die gruppenspezifische Kovarianzmatrix verwendet werden soll.

- **Innerhalb der Gruppen.** Zur Klassifizierung von Fällen wird die gemeinsame Kovarianzmatrix innerhalb der Gruppen verwendet.
- **Gruppenspezifisch.** Für die Klassifizierung werden gruppenspezifische Kovarianzmatrizen verwendet. Da die Klassifizierung auf Diskriminanzfunktionen und nicht auf ursprünglichen Variablen basiert, entspricht diese Option nicht immer der Verwendung einer quadratischen Diskriminanzfunktion.

Diagramme. Die verfügbaren Diagrammoptionen sind "Kombinierte Gruppen", "Gruppenspezifisch" und "Territorien".

- **Kombinierte Gruppen.** Erzeugt ein alle Gruppen umfassendes Streudiagramm der Werte für die ersten beiden Diskriminanzfunktionen. Wenn nur eine Funktion vorliegt, wird stattdessen ein Histogramm angezeigt.
- **Gruppenspezifisch.** Erzeugt gruppenspezifische Streudiagramme der Werte für die ersten beiden Diskriminanzfunktionen. Wenn nur eine Funktion vorliegt, werden stattdessen Histogramme angezeigt.
- **Territorien.** Ein Diagramm der Grenzen, mit denen Fälle auf der Grundlage von Funktionswerten in Gruppen klassifiziert werden. Die Zahlen entsprechen den Gruppen, in die die Fälle klassifiziert wurden. Der Mittelwert jeder Gruppe wird durch einen darin liegenden Stern (*) angezeigt. Dieses Diagramm wird nicht angezeigt, wenn nur eine einzige Diskriminanzfunktion vorliegt.

Diskriminanzanalyse: Speichern

Abbildung 30-7
Dialogfeld "Diskriminanzanalyse: Speichern"

Sie können der aktiven Datendatei neue Variablen hinzufügen. Die verfügbaren Optionen sind "Vorhergesagte Gruppenzugehörigkeit" (eine einzelne Variable), "Wert der Diskriminanzfunktion" (eine Variable für jede Diskriminanzfunktion in der Lösung) und "Wahrscheinlichkeiten der Gruppenzugehörigkeit" unter Berücksichtigung der Werte der Diskriminanzfunktion (eine Variable pro Gruppe).

Des weiteren können Sie Modellinformationen in die angegebene Datei exportieren. SmartScore und SPSS Server (gesondertes Produkt) können anhand dieser Modelldatei die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.

Zusätzliche Funktionen beim Befehl DISCRIMINANT

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Durchführen von mehreren Diskriminanzanalysen (mit einem Befehl) und Festlegen der Reihenfolge, in der die Variablen eingegeben werden (mit dem Unterbefehl `ANALYSIS`).
- Eingeben von a-priori-Wahrscheinlichkeiten für den Klassifikation (mit dem Unterbefehl `PRIORS`).
- Anzeigen von rotierten Mustern und Strukturmatrizen (mit dem Unterbefehl `ROTATE`).
- Begrenzen der Anzahl von extrahierten Diskriminanzfunktionen (mit dem Unterbefehl `FUNCTIONS`).
- Beschränken der Klassifikation auf die Fälle, die für die Analyse ausgewählt (oder nicht ausgewählt) wurden (mit dem Unterbefehl `SELECT`).
- Einlesen und Analysieren der Korrelationsmatrix (mit dem Unterbefehl `MATRIX`).
- Schreiben einer Korrelationsmatrix für die spätere Analyse (mit dem Unterbefehl `MATRIX`).

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
31

Faktorenanalyse

Mit der Faktorenanalyse wird versucht, die zugrunde liegenden Variablen oder **Faktoren** zu bestimmen, welche die Korrelationsmuster innerhalb eines Satzes beobachteter Variablen erklären. Die Faktorenanalyse wird häufig zur Datenreduktion verwendet, indem wenige Faktoren identifiziert werden, welche den größten Teil der in einer großen Anzahl manifester Variablen aufgetretenen Varianz erklären. Die Faktorenanalyse kann auch zum Erzeugen von Hypothesen über kausale Mechanismen oder zum Sichten von Variablen für die anschließende Analyse verwendet werden (zum Beispiel, um vor einer linearen Regressionsanalyse Kollinearität zu erkennen).

Die Prozedur "Faktorenanalyse" bietet ein hohes Maß an Flexibilität:

- Es stehen sieben Methoden der Faktorextraktion zur Verfügung.
- Es sind fünf Rotationsmethoden verfügbar, einschließlich der direkten Oblimin-Methode und Promax-Methode für nicht orthogonale Rotationen.
- Für die Berechnung von Faktorwerten stehen drei Methoden zur Verfügung. Die Werte können für weitere Analysen als Variablen gespeichert werden.

Beispiel. Welche Einstellungen der befragten Personen liegen den gegebenen Antworten bei einer politischen Untersuchung zugrunde? Bei der Untersuchung der Korrelationen zwischen den Themen der Umfrage zeigen sich signifikante Überschneidungen zwischen verschiedenen Untergruppen von Themen. Fragen zu Steuern korrelieren gewöhnlich miteinander, ebenso wie Fragen zum Thema Bundeswehr und so weiter. Mit der Faktorenanalyse können Sie die Anzahl der zugrunde liegenden Faktoren untersuchen und in vielen Fällen die konzeptionelle Bedeutung der Faktoren bestimmen. Zusätzlich können Sie für jeden Fall Faktorwerte berechnen lassen, die sich dann für weiterführende Analysen verwenden lassen. Zum Beispiel könnten Sie ein logistisches Regressionsmodell erstellen, um das Wahlverhalten auf der Grundlage von Faktorwerten vorherzusagen.

Statistiken. Für jede Variable: Anzahl gültiger Fälle, Mittelwert und Standardabweichung. Für jede Faktorenanalyse: Korrelationsmatrix der Variablen mit Signifikanzniveaus, Determinante, Inverse; reproduzierte Korrelationsmatrix mit Anti-Image; Anfangslösung (Kommunalitäten, Eigenwerte und Prozentsatz der erklärten Varianz); Kaiser-Meyer-Olkin-Maß für die Angemessenheit der Stichproben und Bartlett-Test auf Sphärizität; nicht rotierte Lösung mit Faktorladungen, Kommunalität und Eigenwerten; sowie rotierte Lösung mit rotierter Mustermatrix und Transformationsmatrix. Für schiefe Rotationen: rotierte Muster- und Strukturmatrizen; Koeffizientenmatrix der Faktorwerte und Kovarianzmatrix des Faktors. Diagramme: Screeplot von Eigenwerten und Diagramm der Ladungen der ersten zwei oder drei Faktoren.

Daten. Die Variablen müssen auf dem **Intervall-** oder **Verhältnis-**Niveau quantitativ sein. Kategoriale Daten (wie beispielsweise Religion oder Geburtsland) sind für die Faktorenanalyse nicht geeignet. Daten, für welche die Korrelationskoeffizienten nach Pearson sinnvoll berechnet werden können, eignen sich gewöhnlich für eine Faktorenanalyse.

Annahmen. Die Daten sollten für jedes Variablenpaar in einer bivariaten Normalverteilung vorliegen. Beobachtungen müssen unabhängig sein. Im Modell der Faktorenanalyse ist festgelegt, dass Variablen durch gemeinsame Faktoren (die vom Modell geschätzten Faktoren) und eindeutige Faktoren (die sich nicht zwischen den beobachteten Variablen überschneiden) bestimmt sind. Die errechneten Schätzwerte basieren auf der Annahme, dass alle eindeutigen Faktoren weder miteinander noch mit den gemeinsamen Faktoren korrelieren.

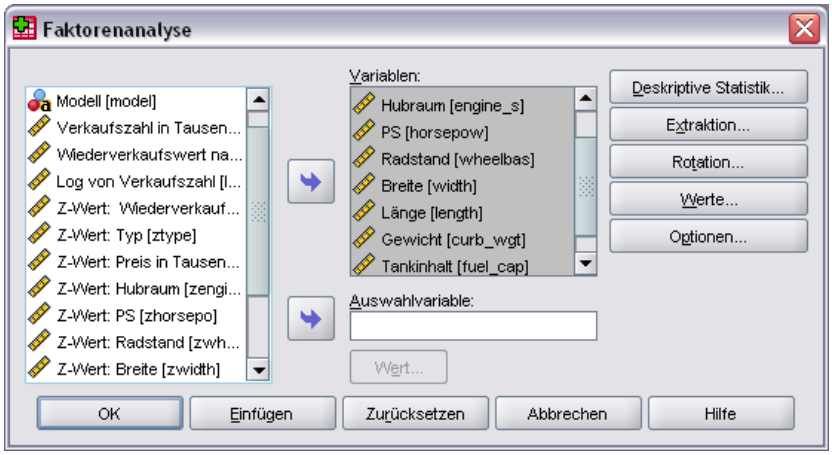
So lassen Sie eine Faktorenanalyse berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Dimensionsreduktion
    Faktorenanalyse...

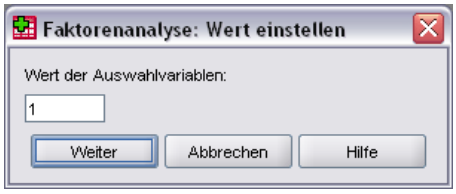
▶ Wählen Sie die Variablen für die Faktorenanalyse aus.

Abbildung 31-1
Dialogfeld "Faktorenanalyse"



Faktorenanalyse: Fälle auswählen

Abbildung 31-2
Dialogfeld "Faktorenanalyse: Wert einstellen"



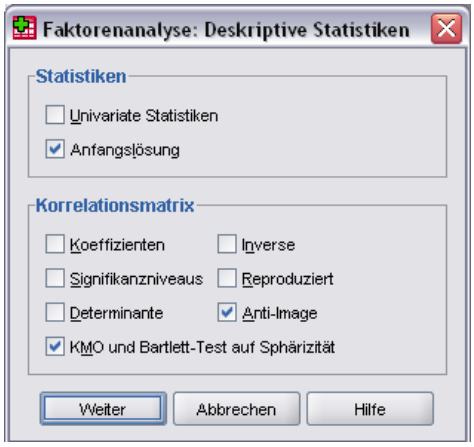
So wählen Sie die Fälle für die Analyse aus:

- Wählen Sie eine Auswahlvariable aus.
- Klicken Sie auf Wert, um eine ganze Zahl als Auswahlvariable einzugeben.

Nur Fälle mit diesem Wert für die Auswahlvariable werden für die Faktorenanalyse verwendet.

Faktorenanalyse: Deskriptive Statistiken

Abbildung 31-3
Dialogfeld "Faktorenanalyse: Deskriptive Statistiken"



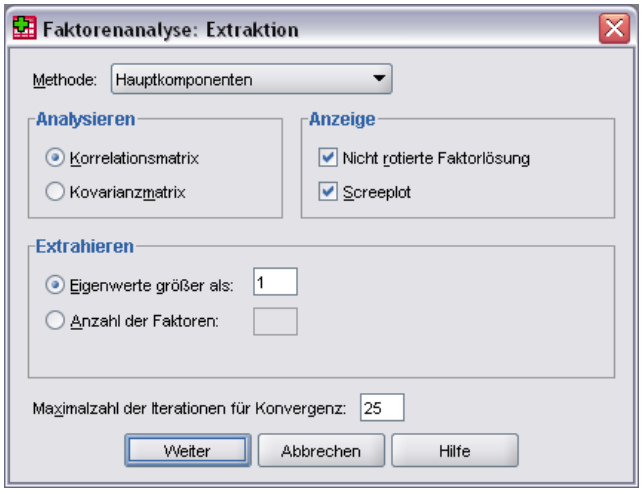
Statistik. Univariate Statistiken enthalten den Mittelwert, die Standardabweichung und die Anzahl gültiger Fälle für jede Variable. Die Anfangslösung zeigt die anfänglichen Kommunalitäten, Eigenwerte und den Prozentwert der erklärten Varianz an.

Korrelationsmatrix. Die verfügbaren Optionen sind Koeffizienten, Signifikanzniveaus, Determinante, Inverse, Reproduziert, Anti-Image sowie KMO und Bartlett-Test auf Sphärizität.

- **KMO und Bartlett-Test auf Sphärizität.** Das Kaiser-Meyer-Olkin-Maß für Angemessenheit der Stichproben überprüft, ob die partiellen Korrelationen zwischen Variablen klein sind. Der Bartlett-Test auf Sphärizität prüft, ob die Korrelationsmatrix eine Einheitsmatrix ist, wobei das Faktorenmodell in diesem Fall ungeeignet wäre.
- **Reproduziert.** Die geschätzte Korrelationsmatrix aus der Faktorlösung. Residuen (Differenz zwischen geschätzten und beobachteten Korrelationen) werden ebenfalls angezeigt.
- **Anti-Image.** Die Anti-Image-Korrelationsmatrix enthält die negativen Werte der partiellen Korrelationskoeffizienten. Die Anti-Image-Kovarianzmatrix enthält die negativen Werte der partiellen Kovarianzen. In einem guten Faktorenmodell sind die meisten außerhalb der Diagonalen liegenden Elemente klein. Das Maß der Stichprobeneignung einer Variablen wird auf der Diagonalen der Anti-Image-Korrelationsmatrix angezeigt.

Faktorenanalyse: Extraktion

Abbildung 31-4
Dialogfeld "Faktorenanalyse: Extraktion"



Methode**.** Hier kann die Methode der Faktorenextraktion festgelegt werden. Folgende Methoden sind verfügbar: Hauptkomponenten, ungewichtete kleinste Quadrate, verallgemeinerte kleinste Quadrate, Maximum Likelihood, Hauptachsen-Faktorenanalyse, Alpha-Faktorisierung und Image-Faktorisierung.

- **Hauptkomponentenanalyse.** Eine Methode zur Faktorextraktion. Sie wird verwendet, um unkorrelierte Linearkombinationen der beobachteten Variablen zu bilden. Die erste Komponente besitzt den größten Varianzanteil. Nachfolgende Komponenten erklären stufenweise kleinere Anteile der Varianz. Sie sind alle miteinander unkorreliert. Die Hauptkomponentenanalyse wird zur Ermittlung der Anfangslösung der Faktorenanalyse verwendet. Sie kann verwendet werden, wenn die Korrelationsmatrix singulär ist.
- **Ungewichtete kleinste Quadrate.** Eine Faktorextraktionsmethode, welche die Summe der quadrierten Differenzen zwischen der beobachteten und der reproduzierten Korrelationsmatrix unter Nichtberücksichtigung der Diagonalen minimiert.
- **Verallgemeinerte Methode der kleinsten Quadrate.** Eine Faktorextraktionsmethode, welche die Summe der quadrierten Differenzen zwischen der beobachteten und der reproduzierten Korrelationsmatrix minimiert. Die Korrelationen werden mit dem inversen Wert der Eindeutigkeit gewichtet, sodass Variablen mit großer Eindeutigkeit schwach und solche mit kleiner Eindeutigkeit stärker gewichtet werden.
- **Maximum-Likelihood-Methode.** Eine Methode für die Faktorextraktion, die Parameterschätzer erzeugt, bei denen die Wahrscheinlichkeit am größten ist, dass sie die beobachtete Korrelationsmatrix erzeugt haben, wenn die Stichprobe aus einer multivariaten Normalverteilung stammt. Die Korrelationen werden durch die inverse Eindeutigkeit der Variablen gewichtet und es wird ein iterativer Algorithmus eingesetzt.
- **Hauptachsen-Faktorenanalyse.** Eine Methode der Faktorextraktion aus der ursprünglichen Korrelationsmatrix, bei der die auf der Diagonalen befindlichen quadrierten Korrelationskoeffizienten als Anfangsschätzer der Kommunalitäten verwendet werden. Diese Faktorladungen werden benutzt, um neue Kommunalitäten zu schätzen, welche die alten

Schätzer auf der Diagonalen ersetzen. Die Iterationen werden so lange fortgesetzt, bis die Änderungen in den Kommunalitäten von einer Iteration zur nächsten das Konvergenzkriterium der Extraktion erfüllen.

- **Alpha.** Eine Methode der Faktorextraktion, welche die Variablen in der Analyse als eine Stichprobe aus einer Grundgesamtheit aller potenziellen Variablen betrachtet. Dies vergrößert die Alpha-Reliabilität der Faktoren.
- **Image-Faktorisierung.** Eine Faktorextraktionsmethode, die von Guttman entwickelt wurde und auf der Imagetheorie basiert. Der gemeinsame Teil einer Variablen – partielles Image genannt – ist als ihre lineare Regression auf die verbleibenden Variablen definiert und nicht als eine Funktion von hypothetischen Faktoren.

Analysieren. Hier können Sie entweder eine Korrelationsmatrix oder eine Kovarianzmatrix festlegen.

- **Korrelationsmatrix.** Diese Funktion ist nützlich, wenn die Variablen in Ihrer Analyse anhand verschiedener Skalen gemessen werden.
- **Kovarianzmatrix.** Diese Funktion ist nützlich, wenn Sie die Faktorenanalyse auf mehrere Gruppen mit unterschiedlichen Varianzen für die einzelnen Variablen anwenden möchten.

Extrahieren. Sie können entweder alle Faktoren, deren Eigenwerte über einem festgelegten Wert liegen, oder eine festgelegte Anzahl von Faktoren beibehalten.

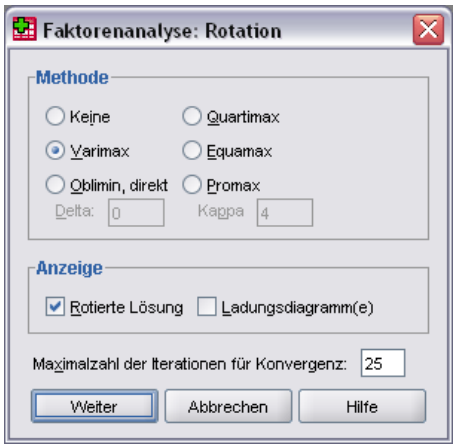
Anzeigen. Hier können Sie die nicht rotierte Faktorlösung und ein Screeplot der Eigenwerte anfordern.

- **Nicht rotierte Faktorlösung.** Zeigt unrotierte Faktorladungen (Faktormustermatrix), Kommunalitäten und Eigenwerte für die Faktorlösung an.
- **Screeplot.** Ein Diagramm der Varianz, die jedem Faktor zugeordnet ist. Es dient dazu, zu bestimmen, wie viele Faktoren beibehalten werden sollen. Normalerweise zeigt das Diagramm einen deutlichen Bruch zwischen der starken Steigung der großen Faktoren und dem graduellen Verlauf der restlichen Faktoren (der "Geröllhalde", engl. "Scree").

Maximalzahl der Iterationen für Konvergenz. Hier können Sie für den Algorithmus eine Maximalzahl von Schritten zum Schätzen der Lösung festlegen.

Faktorenanalyse: Rotation

Abbildung 31-5
Dialogfeld "Faktorenanalyse: Rotation"



Methode. Hier können Sie die Methode der Faktor-Rotation auswählen. Die verfügbaren Methoden sind Varimax, Quartimax, Equamax, Promax oder Oblimin, direkt.

- **Varimax-Rotation.** Eine orthogonale Rotationsmethode, die die Anzahl der Variablen mit hohen Ladungen für jeden Faktor minimiert. Sie vereinfacht die Interpretation der Faktoren.
- **Methode Oblimin, direkt.** Ein Verfahren zur schiefwinkligen (nichtorthogonalen) Rotation. Wenn Delta den Wert 0 annimmt (Standardeinstellung), sind die Ergebnisse am schiefsten. Mit zunehmendem negativem Wert von Delta werden die Faktoren weniger schiefwinklig. Um den Standardwert von 0 zu überschreiben, geben Sie eine Zahl kleiner gleich 0,8 ein.
- **Quartimax-Rotation.** Eine Rotationsmethode, welche die Zahl der Faktoren minimiert, die zum Erklären aller Variablen benötigt werden. Sie vereinfacht die Interpretation der beobachteten Variablen.
- **Equamax-Rotation.** Eine Rotationsmethode, die eine Kombination zwischen der Varimax-Methode (vereinfacht die Faktoren) und der Quartimax-Methode (vereinfacht die Variablen) darstellt. Die Anzahl der Variablen mit hohen Ladungen auf einen Faktor sowie die Anzahl der Faktoren, die benötigt werden, um eine Variable zu erklären, werden minimiert.
- **Promax-Rotation.** Eine schiefe Rotation, bei der Faktoren korreliert sein dürfen. Diese Rotation kann schneller berechnet werden als eine direkte Oblimin-Rotation und ist daher nützlich für große Daten-Sets.

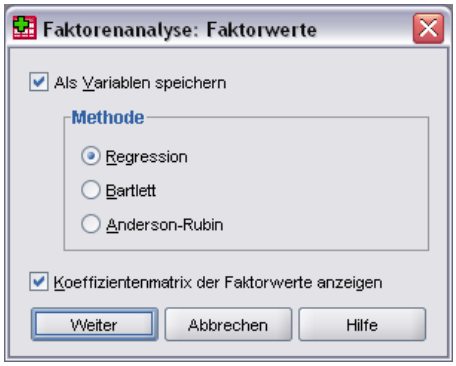
Anzeigen. Hiermit können Sie eine Ausgabe für die rotierte Lösung sowie Ladungsdiagramme für die ersten zwei oder drei Faktoren einbeziehen.

- **Rotierte Lösung.** Um eine rotierte Lösung zu erhalten, muss eine Rotationsmethode ausgewählt sein. Für orthogonale Rotationen werden die rotierte Mustermatrix und Faktortransformationsmatrix angezeigt. Für schiefe Rotationen werden Muster-, Struktur- und Faktorkorrelationsmatrix angezeigt.
- **Diagramm der Faktorladungen.** Dreidimensionales Diagramm der Faktorladungen für die ersten drei Faktoren. Für eine Lösung mit zwei Faktoren wird ein zweidimensionales Diagramm angezeigt. Das Diagramm wird nicht angezeigt, wenn nur ein Faktor extrahiert wird. Auf Wunsch zeigen die Diagramme rotierte Lösungen an.

Maximalzahl der Iterationen für Konvergenz. Hier können Sie eine Maximalzahl von Schritten zum Durchführen der Rotation für den Algorithmus festlegen.

Faktorenanalyse: Faktorwerte

Abbildung 31-6
Dialogfeld "Faktorenanalyse: Faktorwerte"



Als Variablen speichern. Hiermit wird für jeden Faktor in der endgültigen Lösung eine neue Variable erstellt.

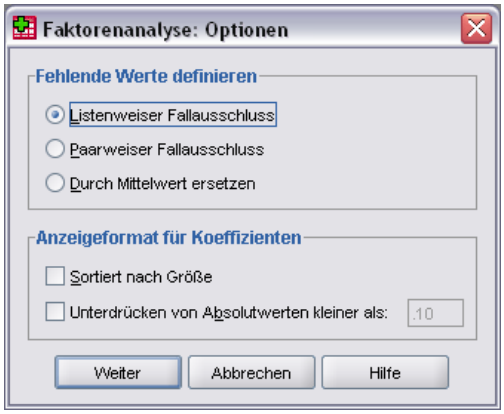
Methode. Als alternative Methoden zur Berechnung der Faktorwerte (Faktor-Scores) sind Regression, Bartlett und Anderson-Rubin.

- **Regressionsmethode.** Eine Methode, um Koeffizienten für Faktorwerte zu schätzen. Die Faktorwerte haben einen Mittelwert von 0 und eine Varianz, die der quadrierten multiplen Korrelation zwischen den geschätzten und den wahren Faktorwerten entspricht. Die Scores können korreliert sein, selbst wenn die Faktoren orthogonal sind.
- **Barlett-Werte.** Eine Methode, um Koeffizienten für Faktorwerte zu schätzen. Die erzeugten Faktorwerte haben einen Mittelwert von 0. Die Quadratsumme der eindeutigen Faktoren über den Variablenbereich wird minimiert.
- **Anderson-Rubin-Methode.** Eine Methode zur Berechnung der Koeffizienten von Faktorwerten; eine Modifizierung der Bartlett-Methode, die die Orthogonalität der geschätzten Faktoren gewährleistet. Die berechneten Werte haben einen Mittelwert von 0 und eine Standardabweichung von 1 und sind unkorreliert.

Koeffizientenmatrix der Faktorwerte anzeigen. Hiermit werden die Koeffizienten angezeigt, mit denen die Variablen multipliziert werden, um Faktorwerte zu erhalten. Hiermit werden auch die Korrelationen zwischen Faktorwerten angezeigt.

Faktorenanalyse: Optionen

Abbildung 31-7
Dialogfeld "Faktorenanalyse: Optionen"



Fehlende Werte. Hier können Sie festlegen, wie fehlende Werte behandelt werden. Es stehen zur Verfügung: "**Listenweiser** Fallausschluss", "**Paarweiser** Fallausschluss" und "Durch Mittelwert ersetzen".

Anzeigeformat für Koeffizienten. Hiermit können Sie Einstellungen für Aspekte der Ausgabematrix vornehmen. Sie können die Koeffizienten nach Größe sortieren lassen und Koeffizienten mit absoluten Werten unterdrücken, die kleiner als der festgelegte Wert sind.

Zusätzliche Funktionen beim Befehl FACTOR

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Angeben von Konvergenzkriterien für die Iteration während der Extraktion und Rotation.
- Angeben von einzelnen rotierten Faktordiagrammen.
- Angeben der Anzahl der zu speichernden Faktorwerte.
- Angeben der Diagonalwerte für die Hauptachsen-Faktorenanalyse.
- Schreiben der Korrelationsmatrizen oder der Faktorladungs-Matrizen auf die Festplatte für eine spätere Analyse.
- Einlesen und Analysieren von Korrelationsmatrizen oder Faktorladungs-Matrizen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
32

Auswählen einer Prozedur zum Durchführen einer Clusteranalyse

Clusteranalysen können mit den Prozeduren “Two-Step-Clusteranalyse”, “Hierarchische Clusteranalyse” oder “Clusterzentrenanalyse” durchgeführt werden. In jeder Prozedur wird ein anderer Algorithmus zum Erstellen von Clustern eingesetzt, und jede Prozedur verfügt über Optionen, die in den jeweils anderen Prozeduren nicht verfügbar sind.

Two-Step-Clusteranalyse. In vielen Fällen ist die Prozedur “Two-Step-Clusteranalyse” die beste Wahl. Sie bietet die folgenden speziellen Funktionen:

- Automatische Auswahl der optimalen Anzahl von Clustern sowie Maße, die bei der Auswahl des Cluster-Modells helfen
- Gleichzeitiges Erstellen von Cluster-Modellen mit kategorialen und stetigen Variablen
- Speichern des Cluster-Modells in einer externen XML-Datei und anschließendem Einlesen dieser Datei und Aktualisieren des Cluster-Modells mit neuen Daten.

Außerdem können von der Prozedur “Two-Step-Clusteranalyse” auch umfangreiche Datendateien analysiert werden.

Hierarchische Clusteranalyse. Die Prozedur “Hierarchische Clusteranalyse” ist auf kleinere Datendateien begrenzt (mehrere Hundert zu gruppierende Objekte), bietet jedoch die folgenden speziellen Funktionen:

- Möglichkeit der Zusammenfassung von Fällen oder Variablen in Clustern
- Funktion zum Berechnen eines Bereichs möglicher Lösungen und zum Speichern der Cluster-Zugehörigkeiten für jede dieser Lösungen
- Verschiedene Methoden zur Clusterbildung, Transformation von Variablen und Messung der Unähnlichkeit zwischen Clustern

Mit der Prozedur “Hierarchische Clusteranalyse” können Intervallvariablen (stetige Variablen), Zählvariablen oder binäre Variablen analysiert werden, wobei alle für die Prozedur ausgewählten Variablen jeweils denselben Typ aufweisen müssen.

Clusterzentrenanalyse. Die Prozedur “Clusterzentrenanalyse” ist auf stetige Daten beschränkt und setzt eine Festlegung der Cluster-Anzahl voraus, bietet jedoch die folgenden speziellen Funktionen:

- Funktion zum Speichern der Distanz vom Clusterzentrum für jedes Objekt
- Funktion zum Einlesen der anfänglichen Clusterzentren aus einer externen SPSS-Datei und zum Speichern der endgültigen Clusterzentren in dieser Datei

Außerdem können von der Prozedur “Clusterzentrenanalyse” auch umfangreiche Datendateien analysiert werden.

Kapitel
33

Two-Step-Clusteranalyse

Bei der Two-Step-Clusteranalyse handelt es sich um eine explorative Prozedur zum Ermitteln von natürlichen Gruppierungen (Clustern) innerhalb eines Daten-Sets, die anderenfalls nicht erkennbar wären. Der von der Prozedur verwendete Algorithmus verfügt über vielfältige nützliche Funktionen, durch die er sich von traditionellen Cluster-Methoden unterscheidet:

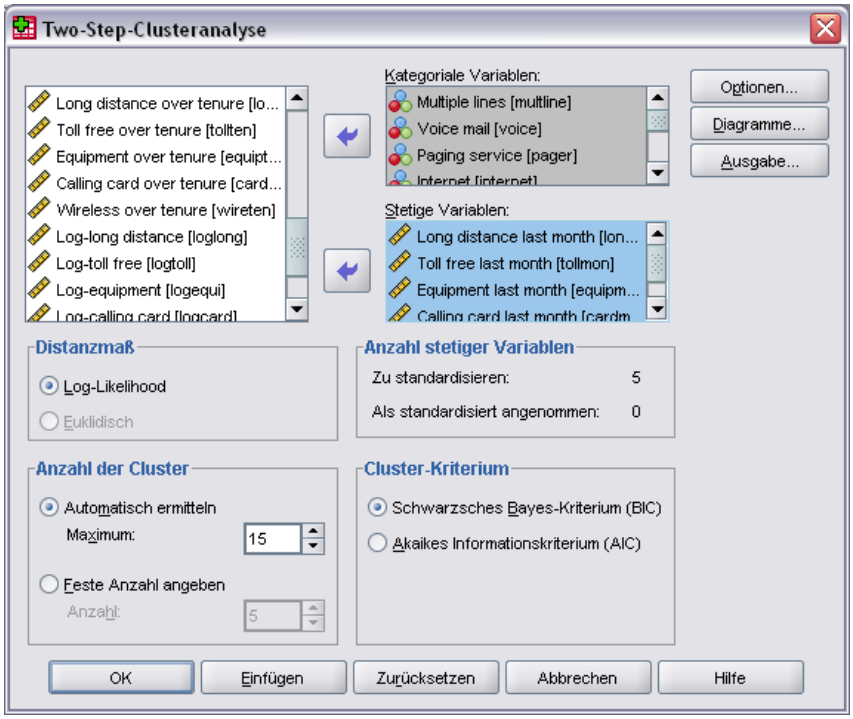
- **Verarbeitung von kategorialen und stetigen Variablen.** Die Annahme der Unabhängigkeit der Variablen ermöglicht eine kombinierte multinomiale Normalverteilung für kategoriale und stetige Variablen.
- **Automatische Auswahl der Cluster-Anzahl.** Durch den Vergleich der Werte eines Modellauswahlkriteriums in verschiedenen Clusteranalysen kann die optimale Anzahl der Cluster von der Prozedur automatisch bestimmt werden.
- **Skalierbarkeit.** Durch das Zusammenfassen der Datensätze in einem Clusterfunktionsbaum (CF-Baum) können mit dem Two-Step-Algorithmus sehr große Datendateien analysiert werden.

Beispiel. In Einzel- und Fachhandel werden Cluster-Methoden regelmäßig auf Daten angewendet, die Kaufgewohnheiten, Geschlecht, Alter und Einkommensniveau der Kundschaft beschreiben. Ziel der Analyse ist eine Ausrichtung der unternehmenseigenen Marketing- und Produktentwicklungsstrategien auf einzelne Konsumentengruppen, um Umsatzsteigerungen und Markentreue zu erreichen.

Statistiken. Mit dieser Prozedur werden Informationskriterien (AIC oder BIC) nach Anzahl der Cluster sowie Cluster-Häufigkeiten und deskriptive Statistiken nach Cluster für die abschließende Clusteranalyse erstellt.

Diagramme. Mit dieser Prozedur werden Balken- und Kreisdiagramme für Cluster-Häufigkeiten sowie Wichtigkeitsdiagramme für Variablen erstellt.

Abbildung 33-1
Dialogfeld "Two-Step-Clusteranalyse"



Distanzmaß. Mit dieser Auswahl legen Sie fest, wie Ähnlichkeiten zwischen zwei Clustern verarbeitet werden.

- **Log-Likelihood.** Mit dem Likelihood-Maß wird eine Wahrscheinlichkeitsverteilung für die Variablen vorgenommen. Bei stetigen Variablen wird von einer Normalverteilung, bei kategorialen Variablen von einer multinomialen Verteilung ausgegangen. Bei allen Variablen wird davon ausgegangen, dass sie unabhängig sind.
- **Euklidisch.** Das Euklidische Maß bezeichnet die "gerade" Distanz zwischen zwei Clustern. Es kann nur dann verwendet werden, wenn es sich bei sämtlichen Variablen um stetige Variablen handelt.

Anzahl der Cluster. Mit dieser Auswahl können Sie angeben, wie die Anzahl der Cluster bestimmt werden soll.

- **Automatisch ermitteln.** Mit dieser Prozedur wird das im Gruppenfeld "Cluster-Kriterium" angegebene Kriterium verwendet, um automatisch die "beste" Anzahl der Cluster zu ermitteln. Sie haben die Möglichkeit, eine positive ganze Zahl für die Höchstzahl der Cluster anzugeben, die von der Prozedur berücksichtigt werden sollen.
- **Feste Anzahl angeben.** Ermöglicht das Festlegen der Anzahl der Cluster für die Analyse. Geben Sie eine positive ganze Zahl ein.

Anzahl stetiger Variablen. Dieses Gruppenfeld enthält eine Zusammenfassung der Standardeinstellungen, die im Dialogfeld "Optionen" für stetige Variablen vorgenommen wurden. Für weitere Informationen siehe Two-Step-Clusteranalyse: Optionen auf S. 402.

Cluster-Kriterium. Mit dieser Auswahl legen Sie fest, wie die Anzahl der Cluster vom automatischen Cluster-Algorithmus bestimmt wird. Angegeben werden kann entweder das Bayes-Informationskriterium (BIC) oder das Akaikes-Informationskriterium (AIC).

Daten. Mit dieser Prozedur können sowohl stetige als auch kategoriale Variablen analysiert werden. Die Fälle bilden dabei die Objekte, die gruppiert werden sollen, während die Variablen die Attribute darstellen, auf deren Grundlage die Gruppierung erfolgt.

Fallreihenfolge. Beachten Sie, dass der Cluster-Funktionsbaum und die endgültige Lösung ggf. von der Reihenfolge der Fälle abhängig sein können. Um die Auswirkungen der Reihenfolge zu minimieren, mischen Sie die Fälle in zufälliger Reihenfolge. Prüfen Sie daher die Stabilität einer bestimmten Lösung, indem Sie verschiedene Lösungen abrufen, bei denen die Fälle in einer unterschiedlichen, zufällig ausgewählten Reihenfolge sortiert sind. In schwierigen Situationen mit äußerst umfangreichen Dateien führen Sie statt dessen mehrere Läufe aus, bei denen eine Stichprobe der Fälle in unterschiedlicher, zufälliger Reihenfolge angeordnet ist.

Annahmen. Das Likelihood-Distanzmaß geht davon aus, dass die Variablen im Clustermodell unabhängig sind. Außerdem wird für stetige Variablen eine Normal- bzw. Gauß-Verteilung und für kategoriale Variablen eine multinomiale Verteilung vorausgesetzt. Empirische interne Tests zeigen, dass die Prozedur wenig anfällig gegenüber Verletzungen hinsichtlich der Unabhängigkeitsannahme und der Verteilungsannahme ist. Dennoch sollten Sie darauf achten, wie genau diese Voraussetzungen erfüllt sind.

Mit der Prozedur Bivariate Korrelationen können Sie die Unabhängigkeit zwischen zwei stetigen Variablen überprüfen. Mit der Prozedur Kreuztabellen können Sie die Unabhängigkeit zwischen zwei kategorialen Variablen überprüfen. Mit der Prozedur Mittelwerte können Sie die Unabhängigkeit zwischen einer stetigen und einer kategorialen Variablen überprüfen. Mit der Prozedur Explorative Datenanalyse prüfen Sie die Normalverteilung einer stetigen Variablen. Mit der Prozedur Chi-Quadrat-Test überprüfen Sie, ob eine kategoriale Variable eine bestimmte multinomiale Verteilung aufweist.

So lassen Sie eine Two-Step-Clusteranalyse berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Klassifizieren
    Two-Step-Clusteranalyse...

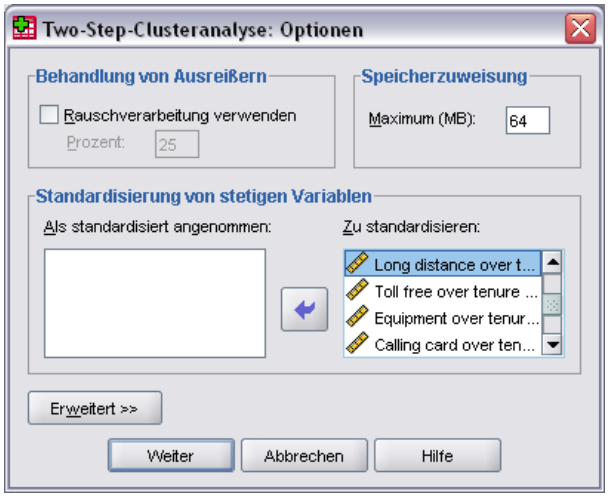
▶ Wählen Sie mindestens eine kategoriale oder stetige Variable aus.

Die folgenden Optionen sind verfügbar:

- Anpassen der Kriterien für die Erstellung der Cluster
- Auswählen der Einstellungen für die Rauschverarbeitung, Speicherzuweisung, Variablenstandardisierung und Eingabe des Clustermodells
- Anfordern von optionalen Tabellen und Diagrammen
- Speichern der Modellergebnisse in der Arbeitsdatei oder in einer externen XML-Datei

Two-Step-Clusteranalyse: Optionen

Abbildung 33-2
Dialogfeld "Two-Step-Clusteranalyse: Optionen"



Behandlung von Ausreißern. Mit diesem Gruppenfeld können Sie Ausreißer während des Füllvorgangs des CF-Baums bei der Clusteranalyse gesondert behandeln. Der CF-Baum ist vollständig, wenn keine weiteren Fälle in einem Blattknoten aufgenommen werden können und kein Blattknoten mehr aufgeteilt werden kann.

- Wenn während des Füllvorgangs des CF-Baums eine Rauschverarbeitung stattfinden soll, wird der CF-Baum neu gebildet, nachdem Fälle von wenig besetzten Blättern auf einem "Rauschblatt" positioniert worden sind. Ein Blatt wird als wenig besetzt betrachtet, wenn es weniger Fälle als den angegebenen Prozentsatz der maximalen Blattgröße enthält. Nach der Neubildung des Baums können gegebenenfalls noch Ausreißer im CF-Baum positioniert werden. Andernfalls werden die Ausreißer verworfen.
- Wenn während des Füllvorgangs des CF-Baums keine Rauschverarbeitung stattfinden soll, wird der Baum unter Verwendung eines größeren Schwellenwerts für die Distanzänderung neu gebildet. Nach der abschließenden Clusteranalyse werden die Werte, die keinem Cluster zugewiesen werden konnten, als Ausreißer bezeichnet. Der Ausreißer-Cluster erhält die Identifikationsnummer –1 und wird nicht in die Auszählung der Anzahl von Clustern aufgenommen.

Speicherzuweisung. In diesem Gruppenfeld können Sie den maximalen Speicherplatz in MB angeben, der vom Cluster-Algorithmus verwenden soll. Wenn der für die Prozedur erforderliche Speicherplatz den maximalen Speicherplatz übersteigt, wird die Festplatte zum Speichern der Daten verwendet, die nicht in den Arbeitsspeicher passen. Geben Sie eine Zahl größer oder gleich 4 ein.

- Den größtmöglichen Wert, den Sie für Ihr System angeben können, erfahren Sie bei Ihrem Systemadministrator.
- Wenn dieser Wert zu niedrig ist, kann die Anzahl der Cluster unter Umständen nicht ordnungsgemäß ermittelt werden.

Variablenstandardisierung. Mit dem Cluster-Algorithmus werden standardisierte stetigen Variablen analysiert. Alle stetigen Variablen, die nicht standardisiert sind, sollten in der Liste "Zu standardisieren" verbleiben. Um Zeit und Verarbeitungsaufwand zu sparen, können Sie alle bereits standardisierten stetigen Variablen in der Liste "Als standardisiert angenommen" auswählen.

Erweiterte Optionen

Verbesserungskriterien für CF-Baum. Die folgenden Einstellungen für den Cluster-Algorithmus gelten insbesondere für den CF-Baum und sollten nur nach sorgfältiger Prüfung geändert werden:

- **Schwellenwert für anfängliche Distanzänderung.** Hierbei handelt es sich um den anfänglichen Schwellenwert, der zum Erstellen des CF-Baums verwendet wird. Wenn das Hinzufügen eines gegebenen Falls zu einem Blatt des CF-Baums eine Dichte unterhalb dieses Schwellenwerts ergibt, wird das Blatt nicht geteilt. Wenn die Dichte den Schwellenwert überschreitet, wird das Blatt geteilt.
- **Höchstzahl Verzweigungen (pro Blattknoten).** Hierbei handelt es sich um die maximale Anzahl an untergeordneten Knoten, über die ein Blattknoten verfügen kann.
- **Maximale Baumtiefe.** Die maximale Anzahl an Ebenen, über die ein CF-Baum verfügen kann.
- **Höchstmögliche Anzahl Knoten.** Gibt die maximale Anzahl an CF-Baumknoten an, die von der Prozedur anhand der Gleichung $(b^{d+1} - 1) / (b - 1)$ potenziell erstellt werden können, wobei b für die Höchstzahl der Verzweigungen und d für die maximale Baumtiefe steht. Beachten Sie, dass ein extrem großer CF-Baum die Systemressourcen stark belastet und somit die Prozedurleistung beinträchtigen kann. Die Mindestanforderung pro Knoten beträgt 16 Bytes.

Aktualisierung des Clustermodells. Mit diesem Gruppenfeld können Sie ein Clustermodell importieren und aktualisieren, das in einer vorangegangenen Analyse erstellt wurde. Die Eingabedatei enthält den CF-Baum im XML-Format. Das Modell wird dann mit den Daten der aktiven Datei aktualisiert. Die Variablennamen müssen im Hauptdialogfeld in der Reihenfolge ausgewählt werden, in der sie in der vorangegangenen Analyse angegeben wurden. Die XML-Datei bleibt unverändert, es sei denn, Sie speichern die neuen Modelldaten unter demselben Dateinamen. Für weitere Informationen siehe Two-Step-Clusteranalyse: Ausgabe auf S. 405.

Bei einer Aktualisierung des Clustermodells werden zur Erstellung des CF-Baums dieselben Optionen verwendet, die für das ursprüngliche Modell gelten. Genauer gesagt werden die Optionen für Distanzmaß, Rauschverarbeitung, Speicherzuweisung und Verbesserungskriterien für den CF-Baum aus dem gespeicherten Modell übernommen, wobei die in den Dialogfeldern für diese Optionen vorgenommenen Einstellungen ignoriert werden.

Anmerkung: Beim Ausführen einer Aktualisierung des Clustermodells wird von der Prozedur vorausgesetzt, dass keiner der ausgewählten Fälle in der Arbeitsdatei für die Erstellung des ursprünglichen Clustermodells verwendet wurde. Außerdem gilt die Annahme, dass die Fälle für die Modellaktualisierung der gleichen Grundgesamtheit entstammen wie die Fälle, die zur Erstellung des ursprünglichen Modells verwendet wurden. Das heißt, es wird angenommen, dass die Mittelwerte und Varianzen der stetigen Variablen sowie die Ebenen der kategorialen Variablen in beiden Fallgruppen identisch sind. Wenn Ihre "neuen" und "alten" Fallgruppen aus heterogenen Grundgesamtheiten stammen, müssen Sie die Two-Step-Clusteranalyse für eine Kombination der beiden Fallgruppen ausführen, um optimale Ergebnisse zu erzielen.

Two-Step-Clusteranalyse: Diagramme

Abbildung 33-3
Dialogfeld "Two-Step-Clusteranalyse: Diagramme"

Prozentdiagramm in Cluster. Hierbei handelt es sich um Diagramme, in denen die Variation der einzelnen Variablen innerhalb eines Clusters angezeigt wird. Für jede kategoriale Variable wird ein gruppiertes Balkendiagramm erstellt, in dem Kategorienhäufigkeiten nach Cluster-ID angezeigt werden. Für jede stetige Variable wird ein Fehlerbalkendiagramm erstellt, in dem Fehlerbalken nach Cluster-ID angezeigt werden.

Gestapeltes Kreisdiagramm. Zeigt ein Kreisdiagramm an, das den Prozentsatz und die Häufigkeit der Beobachtungen innerhalb der einzelnen Cluster darstellt.

Wertigkeitsdiagramme für Variablen. Zeigt zahlreiche unterschiedliche Diagramme an, die die Wichtigkeit der einzelnen Variable in den einzelnen Clustern darstellen. In der Ausgabe werden die einzelnen Variablen nach Wichtigkeitsrang sortiert.

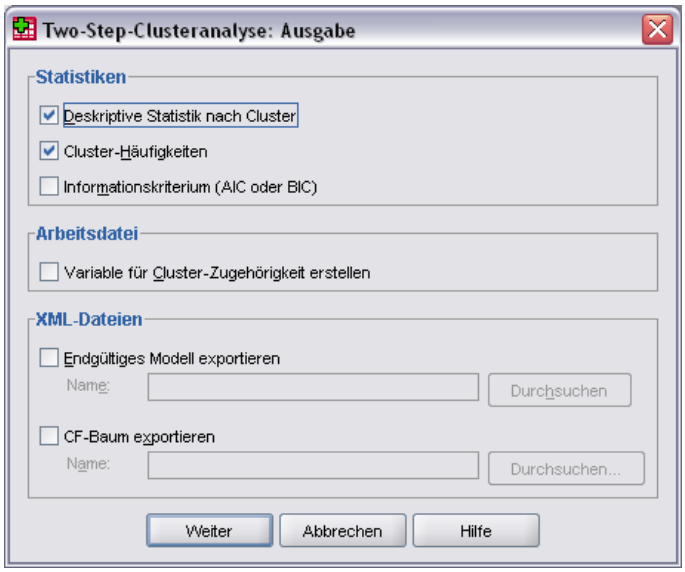
- **Variablenrang.** Mit dieser Option wird festgelegt, für jedes Cluster (Nach Cluster) oder für jede Variable (Nach Variable) ein Diagramm erstellt werden soll.
- **Maß für Wichtigkeit.** Mit dieser Option können Sie festlegen, welches Wichtigkeitsmaß für die Variablen grafisch dargestellt werden soll. Chi-Quadrat oder T-Test der Signifikanz gibt eine Pearson-Chi-Quadrat-Statistik als die Wichtigkeit einer kategorialen Variable und eine *T*-Statistik als Wichtigkeit einer stetigen Variable aus. Signifikanz gibt eins minus *P*-Wert als Test auf Gleichheit der Mittelwerte für eine stetige Variable und die erwartete Häufigkeit im gesamten Daten-Set für eine kategoriale Variable aus.
- **Konfidenzniveau.** Mit dieser Option können Sie das Konfidenzniveau des Tests auf Gleichheit für die Verteilung einer Variablen innerhalb eines Clusters im Vergleich zur Gesamtverteilung der Variablen festlegen. Geben Sie eine Zahl ein, die kleiner als 100 und größer oder gleich 50

ist. Der Wert des Konfidenzniveaus wird als vertikale Linie in den Wichtigkeitsdiagrammen für Variablen dargestellt, wenn die Diagramme für Variablen erstellt werden oder wenn das Signifikanzmaß grafisch dargestellt wird.

- **Nicht signifikante Variablen auslassen.** Variablen, die für das angegebene Konfidenzniveau nicht signifikant sind, werden in den Wichtigkeitsdiagrammen für Variablen nicht angezeigt.

Two-Step-Clusteranalyse: Ausgabe

Abbildung 33-4
Dialogfeld "Two-Step-Clusteranalyse: Ausgabe"



Statistik. In diesem Gruppenfeld können Sie Optionen für die Anzeige von Tabellen mit den Ergebnissen der Clusterananlyse einstellen. Tabellen mit deskriptiven Statistiken und Cluster-Häufigkeiten eignen sich zur Darstellung des endgültigen Clustermodells, während in der Tabelle mit Informationskriterien Ergebnisse für eine Reihe verschiedener Cluster-Lösungen angezeigt werden.

- **Deskriptive Statistik nach Cluster.** Zeigt zwei Tabellen an, die die Variablen in den einzelnen Clustern beschreiben. In der einen Tabelle werden die Mittelwerte und Standardabweichungen der stetigen Variablen nach Cluster erfaßt. In der anderen Tabelle werden die Häufigkeiten der kategorialen Variablen nach Cluster erfaßt.
- **Cluster-Häufigkeiten.** Zeigt eine Tabelle an, in der die Anzahl der Beobachtungen in den einzelnen Clustern erfaßt wird.
- **Informationskriterium (AIC oder BIC).** Zeigt eine Tabelle mit den Werten von AIC oder BIC für eine unterschiedliche Anzahl von Clustern an, je nachdem, welches Kriterium im Hauptdialogfeld ausgewählt wurde. Diese Tabelle wird lediglich dann bereitgestellt, wenn die Anzahl der Cluster automatisch festgelegt wurde. Bei einer festen Anzahl von Clustern wird die Einstellung ignoriert und die Tabelle nicht bereitgestellt.

Arbeitsdatei. Mit diesem Gruppenfeld können Sie Variablen in der Arbeitsdatei speichern.

- **Variable für Cluster-Zugehörigkeit erstellen.** Diese Variable enthält für jeden Fall eine Cluster-Identifikationsnummer. Der Name dieser Variablen lautet *tsc_n*, wobei *n* eine positive ganze Zahl ist, die auf die Ordinalzahl der Arbeitsdatei hinweist, die von dieser Prozedur in einer gegebenen Sitzung gespeichert wurde.

XML-Dateien. Das endgültige Clustermodell und der CF-Baum sind zwei Arten von Ausgabedateien, die als XML-Format exportiert werden können.

- **Endgültiges Modell exportieren.** Das endgültige Clustermodell wird in die angegebene Datei exportiert. SmartScore und SPSS Server (gesondertes Produkt) können anhand dieser Modelldatei die Modellinformationen zu Bewertungszwecken auf andere Datendateien anwenden.
- **CF-Baum exportieren.** Mit dieser Option können Sie den aktuellen Stand des Cluster-Baums speichern und zu einem späteren Zeitpunkt mit neuen Daten aktualisieren.

Kapitel
34

Hierarchische Clusteranalyse

Mit diesem Verfahren wird anhand ausgewählter Merkmale versucht, relativ homogene Fallgruppen oder Variablen zu identifizieren. Dabei wird ein Algorithmus eingesetzt, der für jeden Fall oder für jede Variable, einen separaten Cluster bildet und die Cluster so lange kombiniert, bis nur noch einer zurückbleibt. Sie können einfache Variablen analysieren oder eine Auswahl aus einer Vielfalt von Transformationen zur Standardisierung treffen. Distanz- oder Ähnlichkeitsmaße werden durch die Prozedur "Ähnlichkeiten" erzeugt. Für jeden Schritt werden Statistiken angezeigt, um Sie bei der Auswahl der besten Lösung zu unterstützen.

Beispiel. Können Gruppen von verschiedenen Fernseh-Shows identifiziert werden, die ein ähnliches Publikum ansprechen? Mithilfe der hierarchischen Clusteranalyse können Sie die Fernseh-Shows (Fälle) anhand der Merkmale der Zuschauer in homogene Gruppen (Cluster) aufteilen. Damit lassen sich beispielsweise Marktsegmente identifizieren. Sie können außerdem Städte (Fälle) in homogene Gruppen clustern, sodass vergleichbare Städte zum Testen verschiedener Marketingstrategien ausgewählt werden können.

Statistiken. Zuordnungsübersicht, Distanz- oder Ähnlichkeitsmatrix und Cluster-Zugehörigkeit für eine einzelne Lösung oder einen Bereich von Lösungen. Diagramme: Dendrogramme und Eiszapfendiagramme.

Daten. Bei den Variablen kann es sich um quantitative Daten, binäre Daten oder Häufigkeitsdaten handeln. Die Skalierung der Variablen spielt eine wichtige Rolle. Unterschiede in der Skalierung können sich auf Ihre Cluster-Lösung(en) auswirken. Wenn Ihre Variablen sehr unterschiedlich skaliert sind, eine also beispielsweise in Dollar und die andere in Jahren angegeben wird, empfiehlt sich die Standardisierung. (Die Prozedur "Hierarchische Clusteranalyse" kann dies automatisch durchführen.)

Fallreihenfolge. Wenn gebundene Distanzen oder Ähnlichkeiten in den Eingabedaten vorliegen (oder beim Verbinden in den aktualisierten Clustern auftreten), ist die resultierende Cluster-Lösung ggf. abhängig von der Reihenfolge der Fälle in der Datei. Prüfen Sie daher die Stabilität einer bestimmten Lösung, indem Sie verschiedene Lösungen abrufen, bei denen die Fälle in einer unterschiedlichen, zufällig ausgewählten Reihenfolge sortiert sind.

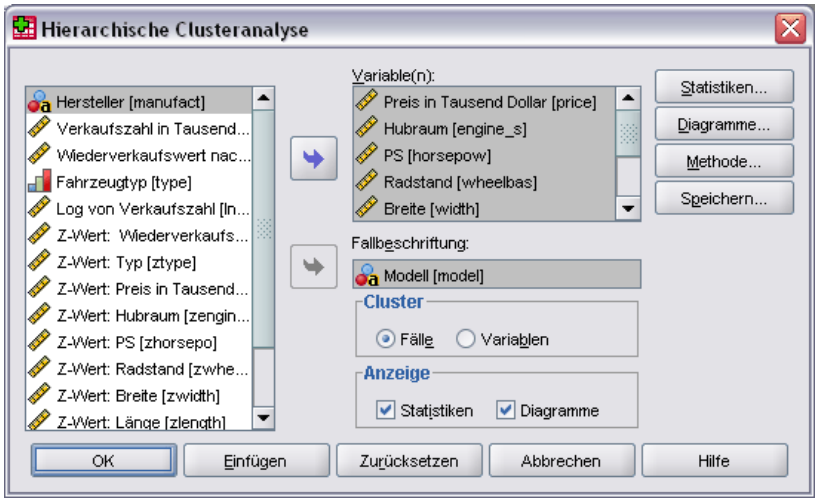
Annahmen. Die verwendeten Distanz- und Ähnlichkeitsmaße müssen für die analysierten Daten geeignet sein. Weitere Informationen zur Auswahl der Distanz- und Ähnlichkeitsmaße finden Sie unter der Prozedur "Ähnlichkeiten". Außerdem sollten Sie alle relevanten Variablen in Ihre Analyse einschließen. Das Weglassen einflußreicher Variablen kann zu irreführenden Lösungen führen. Da es sich bei der hierarchischen Clusteranalyse um eine explorative Methode handelt, sollten die Ergebnisse als vorläufig gelten, bis diese durch eine unabhängige Stichprobe bestätigt werden.

So führen Sie eine hierarchische Clusteranalyse durch:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Klassifizieren
  Hierarchische Cluster…

Abbildung 34-1
Dialogfeld "Hierarchische Clusteranalyse"



- Beim Clustern von Fällen müssen Sie mindestens eine numerische Variable auswählen. Beim Clustern von Variablen müssen Sie mindestens drei numerische Variablen auswählen.

  Sie haben auch die Möglichkeit, eine Variable für die Beschriftung der Fälle auszuwählen.

Hierarchische Clusteranalyse: Methode

Abbildung 34-2
Dialogfeld "Hierarchische Clusteranalyse: Methode"

Cluster-Methode. Verfügbar sind Linkage zwischen den Gruppen, Linkage innerhalb der Gruppen, nächstgelegener Nachbar, entferntester Nachbar, Zentroid-Clustering, Median-Clustering und die Ward-Methode.

Maß. Hiermit können Sie das Distanz- oder Ähnlichkeitsmaß bestimmen, das beim Clustern verwendet wird. Wählen Sie den Typ der Daten sowie das geeignete Distanz- oder Ähnlichkeitsmaß aus.

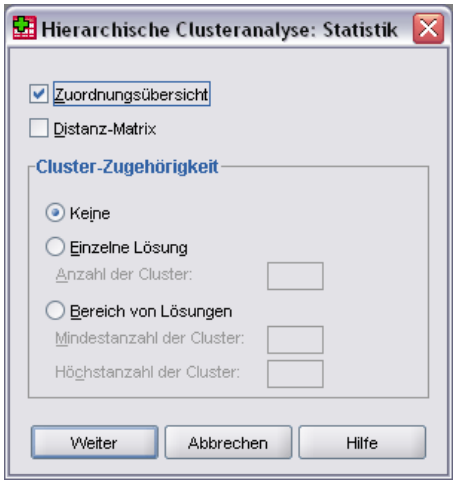
- **Intervall.** Verfügbar sind euklidische Distanz, quadrierte euklidische Distanz, Kosinus, Pearson-Korrelation, Tschebyscheff, Block, Minkowski und die Option Benutzerdefiniert.
- **Häufigkeiten.** Verfügbar sind Chi-Quadratmaß und Phi-Quadratmaß.
- **Binär.** Verfügbar sind euklidische Distanz, quadrierte euklidische Distanz, Größendifferenz, Musterdifferenz, Varianz, Streuung, Form, einfache Übereinstimmung, Phi-4-Punkt-Korrelation, Lambda, Anderberg-*D*, Würfel, Hamann, Jaccard, Kulczynski 1, Kulczynski 2, Distanzmaß nach Lance und Williams, Ochiai, Ähnlichkeitsmaß nach Rogers und Tanimoto, Russel und Rao, Ähnlichkeitsmaße nach Sokal und Sneath 1 bis 5, Yule-*Y* und Yule-*Q*.

Werte transformieren. Hier können Sie festlegen, ob die Datenwerte für Fälle oder Werte vor dem Berechnen von Ähnlichkeiten standardisiert werden (nicht für binäre Daten verfügbar). Die verfügbaren Standardisierungsmethoden sind "*Z*-Scores", "Bereich −1 bis 1", "Bereich 0 bis 1", "Maximale Größe von 1", "Mittelwert 1" und "Standardabweichung 1".

Maße transformieren. Hier können Sie festlegen, ob die durch das Distanzmaß erzeugten Werte transformiert werden. Dies erfolgt, nachdem das Distanzmaß berechnet wurde. Zu den verfügbaren Alternativen zählen Absolutwerte, Ändern des Vorzeichens und Skalieren auf den Bereich 0–1.

Hierarchische Clusteranalyse: Statistik

Abbildung 34-3
Dialogfeld "Hierarchische Clusteranalyse: Statistik"



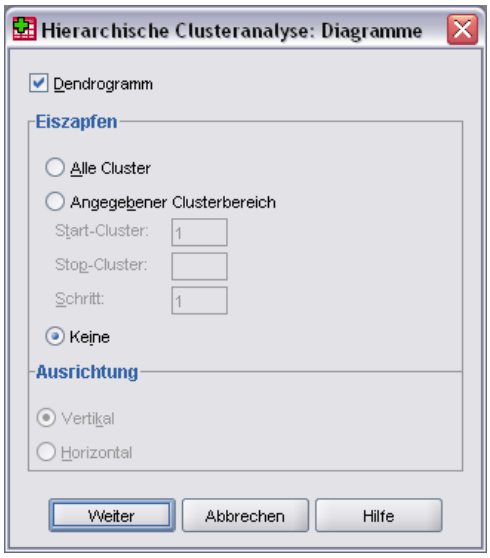
Zuordnungsübersicht. Hier wird folgendes angezeigt: Welche Fälle bzw. Cluster in jedem Schritt kombiniert wurden, die Abstände zwischen den Fällen oder Clustern, die kombiniert werden, und der Cluster-Schritt, in dem ein Fall (oder eine Variable) in den Cluster aufgenommen wurde.

Distanz-Matrix. Zeigt die Distanzen oder Ähnlichkeiten zwischen den Objekten.

Cluster-Zugehörigkeit. Zeigt den Cluster an, dem alle Fälle beim Kombinieren der Cluster in einem oder mehreren Schritten zugeordnet wurden. Die Optionen "Einzelne Lösung" und "Bereich von Lösungen" stehen zur Verfügung.

Hierarchische Clusteranalyse: Diagramme

Abbildung 34-4
Dialogfeld "Hierarchische Clusteranalyse: Diagramme"

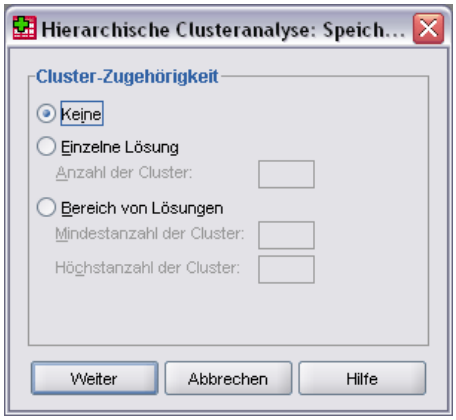


Dendrogramm. Zeigt ein **Dendrogramm** an. Dendrogramme können verwendet werden, um die Dichte der gebildeten Cluster zu bewerten. Sie enthalten Informationen über die angemessene Anzahl der Cluster, die beibehalten werden sollen.

Eiszapfen. Zeigt ein **Eiszapfendiagramm** an, das alle Cluster oder einen bestimmten Bereich von Clustern enthält. Eiszapfendiagramme zeigen an, wie Fälle bei jeder Iteration der Analyse in Clustern zusammengeführt werden. Unter Orientierung können Sie ein vertikales oder horizontales Diagramm auswählen.

Hierarchische Clusteranalyse: Neue Variablen

Abbildung 34-5
Dialogfeld "Hierarchische Clusteranalyse: Neue Variablen speichern"



Cluster-Zugehörigkeit. Hiermit können Sie die Cluster-Zugehörigkeit für eine einzelne Lösung oder einen Bereich von Lösungen speichern. Die gespeicherten Variablen können dann in nachfolgenden Analysen verwendet werden, um andere Differenzen zwischen Gruppen zu untersuchen.

Zusätzliche Funktionen beim Befehl CLUSTER

In der Prozedur "Hierarchische Clusteranalyse" wird die Befehlssyntax von `CLUSTER` verwendet. Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Verwenden mehrerer Cluster-Methoden in einer einzigen Analyse
- Einlesen und Analysieren einer Distanzmatrix
- Schreiben einer Distanzmatrix auf die Festplatte für eine spätere Analyse
- Angeben aller Werte für den Exponenten und die Wurzel im benutzerdefinierten (exponentiellen) Distanzmaß
- Festlegen der Namen für gespeicherte Variablen

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
35

Clusterzentrenanalyse

Diese Prozedur kann relativ homogene Fallgruppen aufgrund ausgewählter Eigenschaften identifizieren, wobei ein Algorithmus verwendet wird, der eine große Anzahl von Fällen verarbeiten kann. Der Algorithmus erfordert jedoch, dass Sie die Anzahl der Cluster festlegen. Wenn Ihnen die anfänglichen Clusterzentren bekannt sind, können Sie diese angeben. Sie können eine der beiden Methoden zur Klassifikation der Fälle auswählen, entweder iteratives Aktualisieren der Clusterzentren oder nur Klassifizieren. Sie können Cluster-Zugehörigkeit, Informationen zur Distanz und endgültige Clusterzentren speichern. Wahlweise können Sie eine Variable festlegen, mit deren Werte fallweise Ausgaben beschriftet werden. Sie können außerdem eine *F*-Statistik zur Varianzanalyse anfordern. Während es sich bei dieser Statistik um eine opportunistische Statistik handelt (mit dieser Prozedur wird versucht, tatsächlich voneinander abweichende Gruppen zu bilden), lassen sich aus der relativen Größe der Statistik Informationen über den Beitrag jeder Variablen zu der Trennung der Gruppen gewinnen.

Beispiel. Wodurch können Gruppen von Fernseh-Shows identifiziert werden, die innerhalb jeder Gruppe ein ähnliches Publikum anziehen? Mit der Clusterzentrenanalyse könnten Sie Fernseh-Shows (Fälle) anhand der Merkmale der Zuschauer in *k* homogene Gruppen clustern. Damit lassen sich beispielsweise Marktsegmente identifizieren. Sie können außerdem Städte (Fälle) in homogene Gruppen clustern, sodass vergleichbare Städte zum Testen verschiedener Marketingstrategien ausgewählt werden können.

Statistiken. Vollständige Lösung: anfängliche Clusterzentren, ANOVA-Tabelle. Jeder Fall: Cluster-Informationen, Distanz vom Clusterzentrum.

Daten. Die Variablen müssen quantitativ sein, entweder auf dem Intervall- oder Verhältnisniveau. Wenn Ihre Variablen binär sind oder Häufigkeiten darstellen, verwenden Sie die Prozedur "Hierarchische Clusteranalyse".

Reihenfolge der Fälle und der anfänglichen Clusterzentren. Der Standardalgorithmus zum Auswählen der anfänglichen Clusterzentren ist nicht invariant bezüglich der Fallreihenfolge. Mit der Option Gleitende Mittelwerte verwenden im Dialogfeld "Iterieren" wird die resultierende Lösung potenziell abhängig von der Reihenfolge der Fälle, unabhängig davon, auf welche Weise die anfänglichen Clusterzentren ausgewählt wurden. Wenn Sie eine dieser Methoden nutzen, prüfen Sie daher die Stabilität einer bestimmten Lösung, indem Sie verschiedene Lösungen abrufen, bei denen die Fälle in einer unterschiedlichen, zufällig ausgewählten Reihenfolge sortiert sind. Wenn Sie anfängliche Clusterzentren angeben und dabei nicht die Option Gleitende Mittelwerte verwenden aktivieren, vermeiden Sie so potentielle Probleme im Zusammenhang mit der Fallreihenfolge. Die Reihenfolge der anfänglichen Clusterzentren kann sich jedoch auf die Lösung auswirken, wenn gebundene Distanzen von Fällen zu Clusterzentren vorliegen. Um die Stabilität einer bestimmten Lösung zu bewerten, können Sie die Ergebnisse von Analysen mit verschiedenen Permutationen der Zentrumsanfangswerte vergleichen.

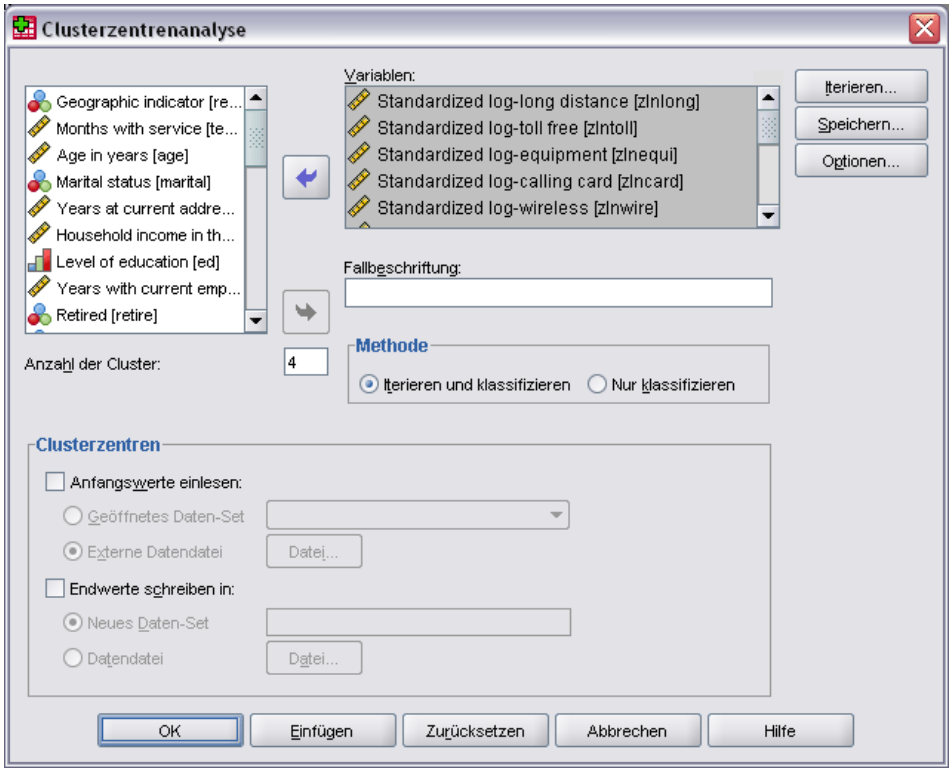
Annahmen. Distanzen werden unter Verwendung des einfachen euklidischen Abstands berechnet. Wenn Sie ein anderes Distanz- oder Ähnlichkeitsmaß verwenden möchten, verwenden Sie die Prozedur "Hierarchische Clusteranalyse". Die Skalierung der Variablen ist eine wichtige Überlegung. Wenn Ihre Variablen auf unterschiedlichen Skalen gemessen wurden (wenn zum Beispiel eine Variable in Dollar und eine andere in Jahren ausgedrückt wird), können die Ergebnisse irreführend sein. In solchen Fällen sollten Sie eine Standardisierung Ihrer Variablen in Betracht ziehen, bevor Sie die Clusterzentrenanalyse durchführen (mit der Prozedur "Deskriptive Statistiken"). Diese Prozedur setzt voraus, dass Sie die passende Anzahl von Clustern ausgewählt und alle relevanten Variablen eingeschlossen haben. Wenn Sie eine ungeeignete Anzahl von Clustern ausgewählt oder wichtige Variablen ausgelassen haben, können Ihre Ergebnisse irreführend sein.

So lassen Sie eine Clusterzentrenanalyse berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Klassifizieren
    Clusterzentrenanalyse…

Abbildung 35-1
Dialogfeld "Clusterzentrenanalyse"



▶ Wählen Sie die Variablen für die Clusteranalyse aus.

▶ Legen Sie die Anzahl der Cluster fest. (Die Anzahl der Cluster muss mindestens 2 betragen und darf nicht größer als die Anzahl der Fälle in der Datendatei sein.)

- Wählen Sie als Methode entweder Iterieren und klassifizieren oder Nur klassifizieren.
- Wählen Sie optional eine Identifizierungsvariable zum Beschriften der Fälle aus.

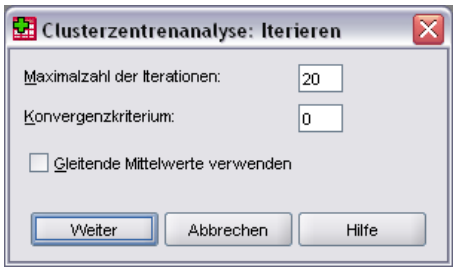
Clusterzentrenanalyse: Effizienz

Der Befehl "Clusterzentrenanalyse" ist in erster Linie deshalb so effizient, weil er nicht die Distanzen zwischen allen Paaren von Fällen berechnet. Dies wird in vielen Algorithmen zum Clustern, auch beim hierarchischen Clustern, durchgeführt.

Für größtmögliche Effizienz nehmen Sie eine Stichprobe von Fällen und bestimmen die Clusterzentren mit der Methode Iterieren und klassifizieren. Wählen Sie Endwerte schreiben in aus. Stellen Sie anschließend die gesamte Datendatei wieder her und wählen Sie als Methode Nur klassifizieren aus. Wählen Sie Anfangswerte einlesen, um die gesamte Datei anhand der aus der Stichprobe geschätzten Clusterzentren zu klassifizieren. Die Daten können in eine Datei oder in ein Daten-Set geschrieben und aus einer Datei oder einem Daten-Set ausgelesen werden. Daten-Sets sind für die anschließende Verwendung in der gleichen Sitzung verfügbar, werden jedoch nicht als Dateien gespeichert, sofern Sie diese nicht ausdrücklich vor dem Beenden der Sitzung speichern. Die Namen von Daten-Sets müssen den Regeln zum Benennen von Variablen entsprechen. Für weitere Informationen siehe Variablennamen in Kapitel 5 auf S. 83.

Clusterzentrenanalyse: Iterieren

Abbildung 35-2
Dialogfeld "Clusterzentrenanalyse: Iterieren"



Anmerkung: Diese Optionen sind nur verfügbar, wenn Sie im Dialogfeld "Clusterzentrenanalyse" die Methode Iterieren und klassifizieren auswählen.

Maximalzahl der Iterationen. Begrenzt die Anzahl der Iterationen im Clusterzentren-Algorithmus. Die Iteration wird nach der vorgegebenen Anzahl der Iterationen beendet, auch wenn das Konvergenzkriterium noch nicht erreicht wurde. Diese Zahl muss zwischen 1 und 999 liegen.

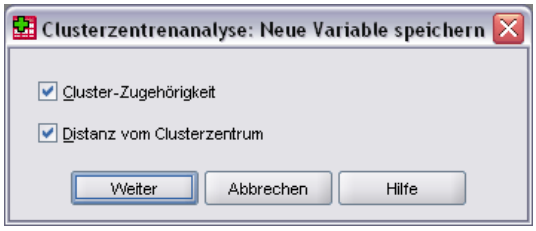
Um den Algorithmus zu verwenden, der beim Befehl "Quick Cluster" in SPSS-Versionen vor Version 5.0 verwendet wurde, setzen Sie Anzahl der Iterationen auf 1.

Konvergenzkriterium. Bestimmt, wann die Iteration beendet ist. Das Konvergenzkriterium gibt einen Anteil der minimalen Distanz zwischen anfänglichen Clusterzentren wieder. Der Wert muss also größer als 0, darf aber nicht größer als 1 sein. Wenn das Kriterium zum Beispiel 0,02 lautet, ist die Iteration beendet, sobald eine vollständige Iteration keines der Clusterzentren um eine Distanz von mehr als 2 % der kleinsten Distanz zwischen beliebigen anfänglichen Clusterzentren bewegt.

Gleitende Mittelwerte verwenden. Mit dieser Funktion können Sie eine Aktualisierung der Clusterzentren veranlassen, nachdem jeder Fall zugeordnet wurde. Wenn Sie diese Option nicht auswählen, werden neue Clusterzentren berechnet, nachdem alle Fälle zugeordnet wurden.

Clusterzentrenanalyse: Neue Variablen

Abbildung 35-3
Dialogfeld "Clusterzentrenanalyse: Neue Variablen speichern"



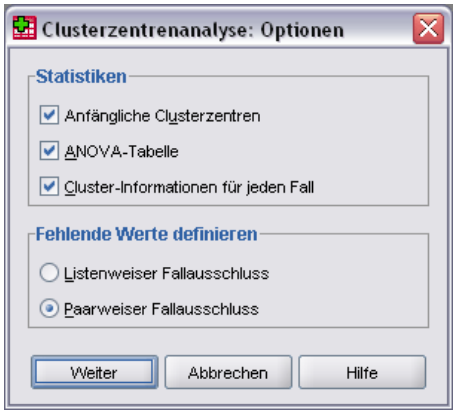
Sie können die Informationen über die Lösung als neue Variablen speichern, um diese in nachfolgenden Analysen zu verwenden:

Cluster-Zugehörigkeit. Erstellt eine neue Variable, welche die endgültige Cluster-Zugehörigkeit für jeden Fall anzeigt. Die Werte der neuen Variablen liegen in einem Bereich von 1 bis zur Anzahl der Cluster.

Distanz vom Clusterzentrum. Erstellt eine neue Variable, welche den euklidischen Abstand zwischen jedem Fall und seinem Klassifikationszentrum anzeigt.

Clusterzentrenanalyse: Optionen

Abbildung 35-4
Dialogfeld "Clusterzentrenanalyse: Optionen"



Statistiken. Sie können die folgenden Statistiken auswählen: anfängliche Clusterzentren, ANOVA-Tabelle und Cluster-Information für jeden Fall.

- **Anfängliche Clusterzentren.** Erster Schätzer der Mittelwerte der Variablen für jeden Cluster. In der Standardeinstellung werden zunächst so viele günstig gelegene Fälle aus den Daten ausgewählt, wie Cluster gebildet werden sollen. Die anfänglichen Clusterzentren werden für eine Ausgangsklassifizierung verwendet und dann aktualisiert.

- **ANOVA-Tabelle.** Zeigt eine Varianzanalysetabelle mit univariaten F-Tests für jede Cluster-Variable an. Die F-Tests haben nur beschreibenden Charakter und die daraus resultierenden Wahrscheinlichkeiten sind nicht zu interpretieren. Die ANOVA-Tabelle wird nicht angezeigt, wenn alle Fälle einem einzigen Cluster zugewiesen werden.
- **Clusterinformationen für die einzelnen Fälle.** Zeigt für jeden Fall die endgültige Clusterzuordnung und den euklidischen Abstand zwischen dem Fall und dem Clusterzentrum, das zur Klassifizierung des Falles verwendet wird. Es werden auch die euklidischen Abstände zwischen den endgültigen Clusterzentren angezeigt.

Fehlende Werte. Die verfügbaren Optionen sind Listenweiser Fallausschluss oder Paarweiser Fallausschluss.

- **Listenweiser Fallausschluss.** Fälle, bei denen Werte einer beliebigen Clustervariable fehlen, werden aus der Analyse ausgeschlossen.
- **Paarweiser Fallausschluss.** Die Fälle werden den Clustern auf der Grundlage der aus allen Variablen mit nichtfehlenden Werten berechneten Distanzen zugewiesen.

Zusätzliche Funktionen beim Befehl QUICK CLUSTER

In der Prozedur "Clusterzentrenanalyse" wird die Befehlssyntax von `QUICK CLUSTER` verwendet. Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Übernehmen der ersten *k* Fälle als anfängliche Clusterzentren. Dadurch wird der üblicherweise für deren Schätzung benötigte Verarbeitungsdurchlauf vermieden.
- Direktes Angeben der anfänglichen Clusterzentren als Teil der Befehlssyntax
- Festlegen der Namen für gespeicherte Variablen

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
36

Nichtparametrische Tests

Mit der Prozedur "Nichtparametrische Tests" stehen Ihnen verschiedene Tests zur Verfügung, bei denen keine Annahmen über die Form der zugrunde liegenden Verteilung benötigt werden.

Chi-Quadrat-Test. Mit diesem Test wird eine Variable nach Kategorien aufgelistet und auf der Grundlage der Differenzen zwischen beobachteten und erwarteten Häufigkeiten eine Chi-Quadrat-Statistik berechnet.

Test auf Binomialverteilung. In diesem Test wird die beobachtete Häufigkeit in jeder Kategorie einer dichotomen Variablen mit den erwarteten Häufigkeiten der binomialen Verteilung verglichen.

Sequenztest. Hiermit können Sie testen, ob zwei Werte einer Variablen in zufälliger Reihenfolge auftreten.

Kolmogorov-Smirnov-Test bei einer Stichprobe. Hierbei wird die beobachtete kumulative Verteilungsfunktion einer Variablen mit einer bestimmten theoretischen Verteilung verglichen. Bei der Verteilung kann es sich um eine Normalverteilung, eine Gleichverteilung, Exponentialverteilung oder um eine Poisson-Verteilung handeln.

Test bei zwei unabhängigen Stichproben. Mit diesem Test können zwei Fallgruppen bei einer Variablen verglichen werden. Dabei stehen die folgenden Tests zur Verfügung: Mann-Whitney-*U*-Test, Kolmogorov-Smirnov-Test bei zwei Stichproben, Test auf Extremreaktionen nach Moses und Sequenzentest nach Wald-Wolfowitz.

Tests bei zwei verbundenen Stichproben. Hiermit können die Verteilungen von zwei Variablen verglichen werden. Dafür stehen der Wilcoxon-Test, der Vorzeichentest und der McNemar-Test zur Verfügung.

Test bei mehreren unabhängigen Stichproben. Hiermit können Sie zwei oder mehrere Fallgruppen bei einer Variablen vergleichen. Dafür stehen der Kruskal-Wallis-H-Test, der Mediantest und der Jonckheere-Terpstra-Test zur Verfügung.

Tests bei mehreren verbundenen Stichproben. Hiermit können Sie die Verteilungen von zwei oder mehr Variablen vergleichen. Dafür stehen der Friedman-Test, Kendall-*W* und Cochrans *Q*-Test zur Verfügung.

Bei allen oben aufgeführten Tests können Quartile, Mittelwert, Standardabweichung, Minimum, Maximum und die Anzahl nichtfehlender Fälle berechnet werden.

Chi-Quadrat-Test

Mit der Prozedur "Chi-Quadrat-Test" können Sie eine Variable nach Kategorien auflisten und eine Chi-Quadrat-Statistik berechnen lassen. Bei diesem Anpassungstest werden die beobachteten und erwarteten Häufigkeiten in allen Kategorien miteinander verglichen. Dadurch wird überprüft, ob

entweder alle Kategorien den gleichen Anteil an Werten enthalten oder ob jede Kategorie jeweils einen vom Benutzer festgelegten Anteil an Werten enthält.

Beispiele. Mithilfe des Chi-Quadrat-Tests können Sie bestimmen, ob in einer Tüte mit Gummibärchen die gleiche Anzahl an weißen, grünen, orangefarbenen, roten und gelben Gummibärchen vorhanden sind. Sie können auch prüfen, ob eine Tüte 30% weiße, 17% grüne, 23% orangefarbene, 15% rote und 15% gelbe Gummibärchen enthält.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum und Quartile. Die Anzahl und der Prozentsatz nichtfehlender und fehlender Fälle, die Anzahl der für jede Kategorie beobachteten und erwarteten Fälle, Residuen und die Chi-Quadrat-Statistik.

Daten. Verwenden Sie geordnete oder nichtgeordnete numerische kategoriale Variablen (nominales oder ordinales Niveau der Messwerte). Verwenden Sie zum Umwandeln von String-Variablen in numerische Variablen den Befehl "Automatisch umkodieren" im Menü "Transformieren".

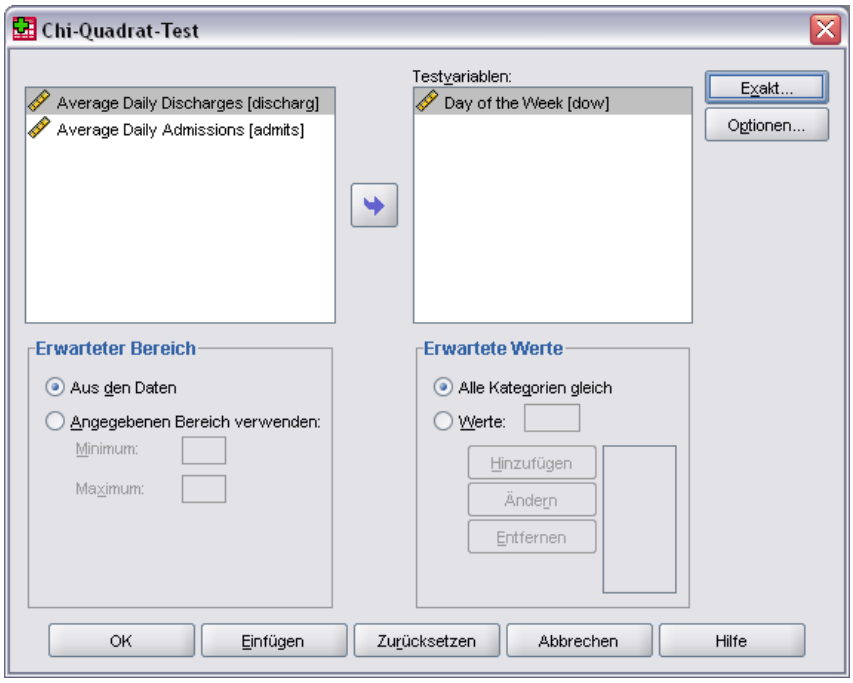
Annahmen. Nichtparametrische Tests erfordern keine Annahmen über die Form der zugrunde liegenden Verteilung. Die Daten werden als zufällige Stichprobe betrachtet. Die erwartete Häufigkeit in jeder Kategorie muss mindestens 1 betragen. Bei höchstens 20% der Kategorien darf die erwartete Häufigkeit unter 5 liegen.

So lassen Sie einen Chi-Quadrat-Test berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    Chi-Quadrat...

Abbildung 36-1
Dialogfeld "Chi-Quadrat-Test"



- Wählen Sie mindestens eine Testvariable aus. Mit jeder Variablen wird ein separater Test erzeugt.
- Wenn Sie auf Optionen klicken, können Sie deskriptive Statistiken und Quartile abrufen sowie festlegen, wie fehlende Werte verarbeitet werden.

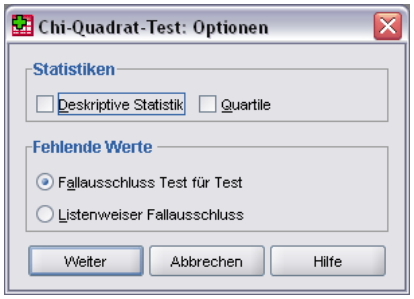
Chi-Quadrat-Test: erwarteter Bereich und erwartete Werte

Erwarteter Bereich. In der Standardeinstellung wird jeder einzelne Wert einer Variablen als eine Kategorie definiert. Zum Aufstellen von Kategorien in einem bestimmten Bereich wählen Sie AngegebenenBereich verwenden und geben Sie für die obere und die untere Grenze jeweils einen ganzzahligen Wert an. Für jeden ganzzahligen Wert in dem eingeschlossenen Bereich wird eine Kategorie aufgestellt, wobei Fälle mit Werten außerhalb der angegebenen Grenzen ausgeschlossen werden. Wenn Sie zum Beispiel für das Minimum den Wert 1 und für das Maximum den Wert 4 angeben, werden für den Chi-Quadrat-Test nur die Werte von 1 bis 4 verwendet.

Erwartete Werte. In der Standardeinstellung sind die erwarteten Werte für alle Kategorien gleich. Die erwarteten Anteile der Kategorien können vom Benutzer festgelegt werden. Wählen Sie Werte aus. Geben Sie für jede Kategorie der Testvariablen einen Wert größer als 0 ein und klicken Sie dann auf Hinzufügen. Jeder neu eingegebene Wert wird am Ende der Werteliste angezeigt. Die Reihenfolge der Werte ist von Bedeutung. Sie entspricht der aufsteigenden Folge der Kategoriewerte für die Testvariable. Der erste Wert in der Liste entspricht dem niedrigsten Gruppenwert der Testvariablen, der letzte Wert entspricht dem höchsten Wert. Die Elemente der Werteliste werden summiert. Anschließend wird jeder Wert durch diese Summe dividiert, um den Anteil der in der entsprechenden Kategorie erwarteten Fälle zu berechnen. So ergibt eine Werteliste mit 3, 4, 5 und 4 beispielsweise die erwarteten Anteile 3/16, 4/16, 5/16 und 4/16.

Chi-Quadrat-Test: Optionen

Abbildung 36-2
Dialogfeld "Chi-Quadrat-Test: Optionen"



Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Bei dieser Option werden Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle angezeigt.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluss Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (Chi-Quadrat-Test)

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Mit dem Unterbefehl `CHISQUARE` können verschiedene Minimal- und Maximalwerte sowie erwartete Häufigkeiten für verschiedene Variablen angegeben werden.
- Mit dem Unterbefehl `EXPECTED` kann eine Variable bei verschiedenen erwarteten Häufigkeiten getestet werden oder es können verschiedene Bereiche verwendet werden.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Test auf Binomialverteilung

Mit der Prozedur "Test auf Binomialverteilung" können Sie die beobachteten Häufigkeiten der beiden Kategorien einer dichotomen Variablen mit den Häufigkeiten vergleichen, die unter einer Binomialverteilung mit einem angegebenen Wahrscheinlichkeitsparameter zu erwarten sind. In der Standardeinstellung ist der Wahrscheinlichkeitsparameter für beide Gruppen auf 0,5 gesetzt. Zum Ändern der Wahrscheinlichkeiten können Sie einen Testanteil für die erste Gruppe angeben. Die Wahrscheinlichkeit für die zweite Gruppe beträgt 1 minus der für die erste Gruppe angegebenen Wahrscheinlichkeit.

Beispiel. Wenn Sie eine Münze werfen, ist die Wahrscheinlichkeit, dass diese mit dem Kopf nach oben zu liegen kommt, gleich 1/2. Auf der Grundlage dieser Hypothese wird nun eine Münze 40mal geworfen, wobei die Ergebnisse aufgezeichnet werden (Kopf oder Zahl). Der Test auf Binomialverteilung könnte dann beispielsweise ergeben, dass 3/4 der Würfe "Kopf" waren und das beobachtete Signifikanzniveau gering ist (0,0027). Diese Ergebnisse zeigen an, dass die Wahrscheinlichkeit für "Kopf" nicht 1/2 beträgt und die Münze somit wahrscheinlich manipuliert ist.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile.

Daten. Die getesteten Variablen müssen numerisch und dichotom sein. Verwenden Sie zum Umwandeln von String-Variablen in numerische Variablen den Befehl "Automatisch umkodieren" im Menü "Transformieren". **Dichotome Variablen** sind Variablen, die nur zwei mögliche Werte annehmen können: *ja* oder *nein*, *wahr* oder *falsch*, 0 oder 1 usw. Wenn die Varaiblen nicht dichotom sind, müssen Sie einen Trennwert angeben. Durch den Trennwert werden Fälle mit Werten über dem Trennwert einer Gruppe und alle anderen Fälle einer anderen Gruppe zugeordnet.

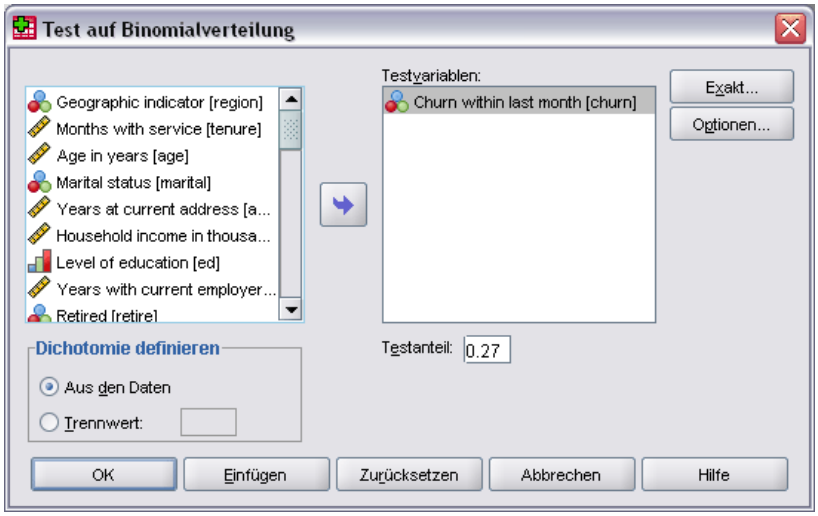
Annahmen. Nichtparametrische Tests erfordern keine Annahmen über die Form der zugrunde liegenden Verteilung. Die Daten werden als zufällige Stichprobe betrachtet.

So lassen Sie einen Test auf Binomialverteilung berechnen:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Nichtparametrische Tests
  Binomial...

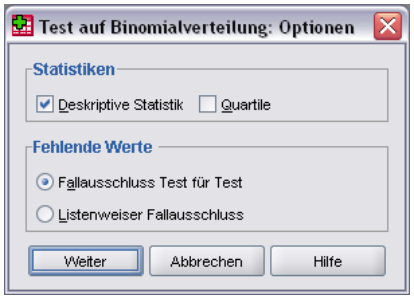
Abbildung 36-3
Dialogfeld "Test auf Binomialverteilung"



- Wählen Sie mindestens eine numerische Testvariable.
- Wenn Sie auf Optionen klicken, können Sie deskriptive Statistiken und Quartile abrufen sowie festlegen, wie fehlende Werte verarbeitet werden.

Optionen für den Test auf Binomialverteilung

Abbildung 36-4
Dialogfeld "Test auf Binomialverteilung: Optionen"



Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Bei dieser Option werden Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle angezeigt.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluss Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für eine beliebige getestete Variable werden von allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (Test auf Binomialverteilung)

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Mit dem Unterbefehl `BINOMIAL` können bestimmte Gruppen ausgewählt und andere Gruppen ausgeschlossen werden, wenn eine Variable über mehr als zwei Kategorien verfügt.
- Mit dem Unterbefehl `BINOMIAL` können verschiedene Trennwerte oder Wahrscheinlichkeiten für verschiedene Variablen angeben werden.
- Mit dem Unterbefehl `EXPECTED` kann dieselbe Variable bei verschiedenen Trennwerten oder Wahrscheinlichkeiten getestet werden.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Sequenzentest

Mit der Prozedur “Sequenzentest” können Sie testen, ob zwei Werte einer Variablen in zufälliger Reihenfolge auftreten. Eine Sequenz ist eine Folge von gleichen Beobachtungen. Eine Stichprobe mit zu vielen oder zu wenigen Sequenzen legt nahe, dass die Stichprobe nicht zufällig ist.

Beispiele. Es werden 20 Personen befragt, ob sie ein bestimmtes Produkt kaufen würden. Die angenommene zufällige Auswahl der Stichprobe wäre ernsthaft zu bezweifeln, wenn alle 20 Personen demselben Geschlecht angehören würden. Mit dem Sequenzentest kann bestimmt werden, ob die Stichprobe zufällig entnommen wurde.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile.

Daten. Die Variablen müssen numerisch sein. Verwenden Sie zum Umwandeln von String-Variablen in numerische Variablen den Befehl “Automatisch umkodieren” im Menü “Transformieren”.

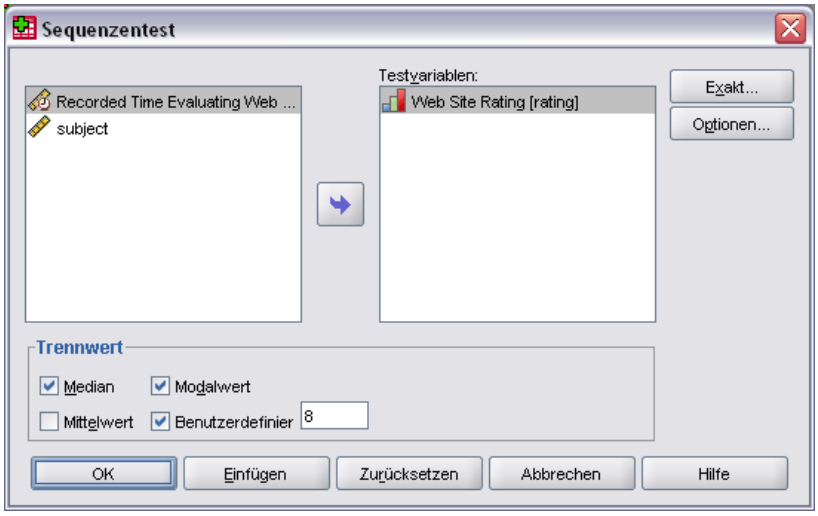
Annahmen. Nichtparametrische Tests erfordern keine Annahmen über die Form der zugrunde liegenden Verteilung. Verwenden Sie Stichproben aus stetigen Wahrscheinlichkeitsverteilungen.

So lassen Sie einen Sequenzentest berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    Sequenzen...

Abbildung 36-5
Hinzufügen eines benutzerdefinierten Trennwerts



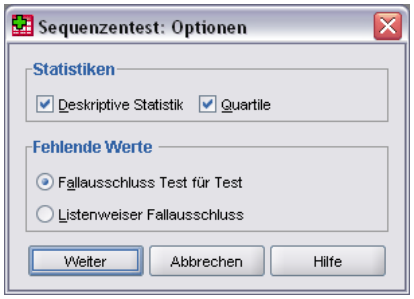
- Wählen Sie mindestens eine numerische Testvariable.
- Wenn Sie auf Optionen klicken, können Sie deskriptive Statistiken und Quartile abrufen sowie festlegen, wie fehlende Werte verarbeitet werden.

Sequenzentest: Trennwert

Trennwert. Hier wird ein Trennwert zum Dichotomisieren der gewählten Variablen angegeben. Sie können den beobachteten Mittelwert, den Median, den Modalwert oder einen angegebenen Wert als Trennwert wählen. Fälle mit Werten kleiner als der Trennwert werden einer Gruppe, Fälle mit Werten größer oder gleich dem Trennwert einer anderen Gruppe zugeordnet. Für jeden gewählten Trennwert wird ein Test ausgeführt.

Sequenzentest: Optionen

Abbildung 36-6
Dialogfeld "Sequenzentest: Optionen"



Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Bei dieser Option werden Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle angezeigt.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluss Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (Sequenzentest)

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Mit dem Unterbefehl `RUNS` können verschiedene Trennwerte für verschiedene Variablen angegeben werden.
- Mit dem Unterbefehl `RUNS` kann dieselbe Variable mit verschiedenen benutzerdefinierten Trennwerten getestet werden.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kolmogorov-Smirnov-Test bei einer Stichprobe

Mit dem Kolmogorov-Smirnov-Test bei einer Stichprobe (Anpassungstest) wird die beobachtete kumulative Verteilungsfunktion für eine Variable mit einer festgelegten theoretischen Verteilung verglichen, die eine Normalverteilung, eine Gleichverteilung, eine Poisson-Verteilung oder Exponentialverteilung sein kann. Das Kolmogorov-Smirnov-*Z* wird aus der größten Differenz (in Absolutwerten) zwischen beobachteten und theoretischen kumulativen Verteilungsfunktionen berechnet. Mit diesem Test für die Güte der Anpassung wird getestet, ob die Beobachtung wahrscheinlich aus der angegebenen Verteilung stammt.

Beispiel. Für viele parametrische Tests sind normalverteilte Variablen erforderlich. Mit dem Kolmogorov-Smirnov-Anpassungstest kann getestet werden, ob eine Variable, zum Beispiel *Einkommen*, normalverteilt ist.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile.

Daten. Die Variablen müssen auf Intervall- oder Verhältnis-Messniveau quantitativ sein.

Annahmen. Für den Kolmogorov-Smirnov-Test wird angenommen, dass die Parameter der zu testenden Verteilung im voraus angegeben wurden. Mit dieser Prozedur werden die Parameter aus der Stichprobe geschätzt. Der Mittelwert und die Standardabweichung der Stichprobe sind die Parameter für eine Normalverteilung. Minimum und Maximum der Stichprobe definieren die Spannweite der Gleichverteilung, und der Mittelwert der Stichprobe ist der Parameter für die Poisson-Verteilung sowie der Parameter für die Exponentialverteilung. Die Stärke des Tests, Abweichungen von der hypothetischen Verteilung zu erkennen, kann dabei deutlich verringert werden. Wenn Sie einen Test gegen eine Normalverteilung mit geschätzten Parametern

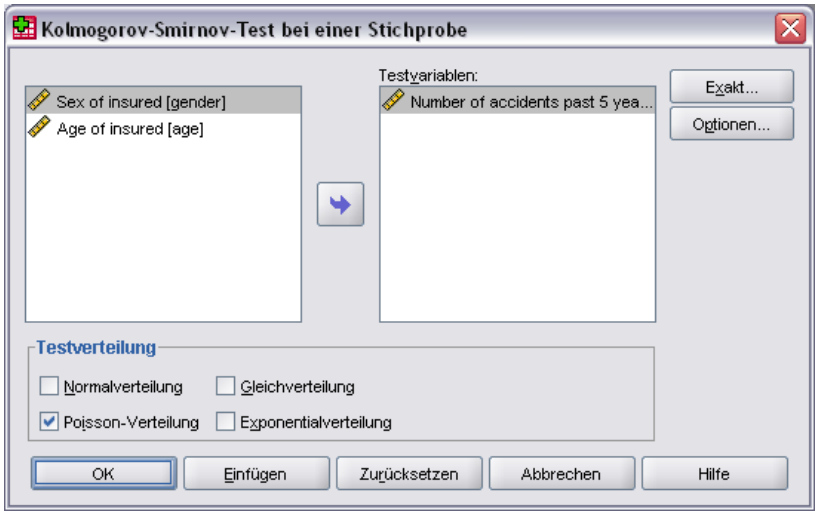
durchführen möchten, sollten Sie den Kolmogorov-Smirnov-Test mit der Korrektur nach Lilliefors (in der Prozedur "Explorative Datenanalyse") in Betracht ziehen.

So berechnen Sie einen Kolmogorov-Smirnov-Anpassungstest:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
Nichtparametrische Tests
K-S bei einer Stichprobe…

Abbildung 36-7
Dialogfeld "Kolmogorov-Smirnov-Test bei einer Stichprobe"

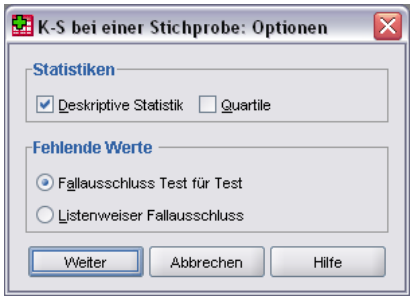


- Wählen Sie mindestens eine numerische Testvariable. Mit jeder Variablen wird ein separater Test erzeugt.

- Wenn Sie auf Optionen klicken, können Sie deskriptive Statistiken und Quartile abrufen sowie festlegen, wie fehlende Werte verarbeitet werden.

K-S bei einer Stichprobe: Optionen

Abbildung 36-8
Dialogfeld "K-S bei einer Stichprobe: Optionen"



Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Bei dieser Option werden Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle angezeigt.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluss Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (Kolmogorov-Smirnov-Anpassungstest)

Mit der Befehlssyntax-Sprache können Sie auch die Parameter der zu testenden Verteilung angeben (mit dem Unterbefehl `K-S`).

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Tests bei zwei unabhängigen Stichproben

Die Prozedur "Test bei zwei unabhängigen Stichproben" vergleicht zwei Gruppen von Fällen von einer Variablen.

Beispiel. Es wurden neue Zahnspangen entwickelt, die bequemer sein sollen, besser aussehen und zu einem schnelleren Erfolg beim Richten der Zähne führen sollen. Um festzustellen, ob die neuen Spangen so lange wie die alten getragen werden müssen, wurden willkürlich 10 Kinder zum Tragen der alten Zahnspangen und weitere 10 Kinder zum Tragen der neuen Spangen ausgewählt. Anhand des Mann-Whitney-*U*-Tests stellen Sie eventuell fest, dass die neuen Spangen im Durchschnitt nicht so lange wie die alten Spangen getragen werden mussten.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile. Tests: Mann-Whitney-*U*-Test, Extremreaktionen nach Moses, Kolmogorov-Smirnov-*Z*-Test, Sequenztest nach Wald-Wolfowitz.

Daten. Verwenden Sie numerische Variablen, die geordnet werden können.

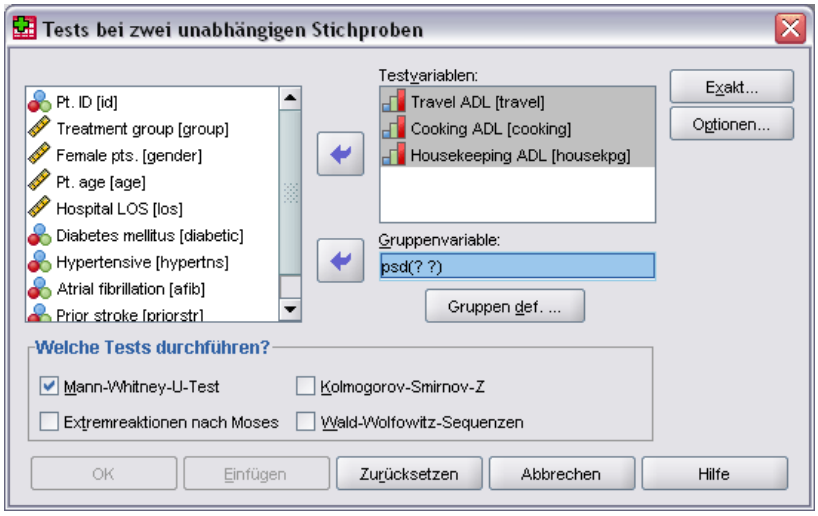
Annahmen. Verwenden Sie unabhängige Zufallsstichproben. Der Mann-Whitney-*U*-Test erfordert, dass sich die beiden getesteten Stichproben in ihrer Form ähneln.

So lassen Sie Tests bei zwei unabhängigen Stichproben berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    Zwei unabhängige Stichproben...

Abbildung 36-9
Dialogfeld "Tests bei zwei unabhängigen Stichproben"



- Wählen Sie mindestens eine numerische Variable aus.
- Wählen Sie eine Gruppenvariable aus und klicken Sie auf Gruppen definieren, um die Datei in zwei Gruppen oder Stichproben aufzuteilen.

Typen von Tests bei zwei unabhängigen Stichproben

Welche Tests durchführen? Mithilfe von vier Tests können Sie überprüfen, ob zwei unabhängige Stichproben (Gruppen) aus derselben Grundgesamtheit stammen.

Der **Mann-Whitney-U-Test**ist der am häufigsten verwendete Test bei zwei unabhängigen Stichproben. Er ist äquivalent zum Wilcoxon-Rangsummentest und dem Kruskal-Wallis-Test für zwei Gruppen. Mit dem Mann-Whitney-U-Test wird überprüft, ob zwei beprobte Grundgesamtheiten die gleiche Lage besitzen. Die Beobachtungen aus beiden Gruppen werden kombiniert und in eine gemeinsame Reihenfolge gebracht, wobei im Falle von Rangbindungen der durchschnittliche Rang vergeben wird. Die Anzahl der Bindungen sollte im Verhältnis zur Gesamtanzahl der Beobachtungen klein sein. Wenn die Grundgesamtheiten in der Lage identisch sind, sollten die Ränge zufällig zwischen den beiden Stichproben gemischt werden. Im Test wird berechnet, wie oft ein Wert aus Gruppe 1 einem Wert aus Gruppe 2 und wie oft ein Wert aus Gruppe 2 einem Wert aus Gruppe 1 vorangeht. Die Mann-Whithney-*U*-Statistik ist die kleinere dieser beiden Zahlen. Die ebenfalls angezeigte Wilcoxon-Rangsummen-*W*-Statistik ist die kleinere der beiden Rangsummen. Wenn beide Stichproben die gleiche Anzahl von Beobachtungen aufweisen, ist *W* die Rangsumme der zuerst benannten Gruppe im Dialogfeld "Zwei unabhängige Stichproben: Gruppen definieren".

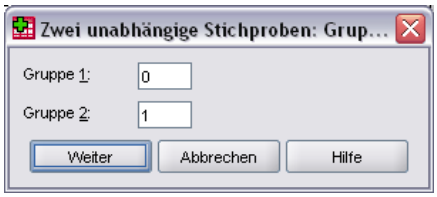
Der **Kolmogorov-Smirnov-Z-Test** und der **Sequenztest nach Wald-Wolfowitz** stellen eher allgemeine Tests dar, die sowohl Unterschiede in den Lagen als auch in den Formen der Verteilungen erkennen. Der Test nach Kolmogorov-Smirnov arbeitet auf der Grundlage der maximalen absoluten Differenz zwischen den beobachteten kumulativen Verteilungsfunktionen für beide Stichproben. Wenn diese Differenz signifikant groß ist, werden die beiden Verteilungen als verschieden betrachtet. Der Sequenztest nach Wald-Wolfowitz kombiniert die Beobachtungen

aus beiden Gruppen und ordnet ihnen einen Rang zu. Wenn die beiden Stichproben aus derselben Grundgesamtheit stammen, müssen die beiden Gruppen in der Rangverteilung zufällig gestreut sein.

Der **Test "Extremreaktionen nach Moses"** setzt voraus, dass die experimentelle Variable einige Subjekte in der einen Richtung und andere Subjekte in der entgegengesetzten Richtung beeinflußt. In diesem Test wird auf extreme Antworten im Vergleich zu einer Kontrollgruppe geprüft. Dieser Test konzentriert sich auf die Spannweite der Kontrollgruppe und ist ein Maß dafür, wie stark die Spannweite durch die extremen Werte in der experimentellen Gruppe beeinflußt wird, wenn sie mit der Kontrollgruppe verbunden werden. Die Kontrollgruppe wird durch den Wert der Gruppe 1 im Dialogfeld "Zwei unabhängige Stichproben: Gruppen definieren" bestimmt. Die Beobachtungen aus beiden Gruppen werden kombiniert und einem Rang zugeordnet. Die Spanne der Kontrollgruppe wird als die Differenz zwischen den Rängen der größten und kleinsten Werte in der Kontrollgruppe plus 1 berechnet. Da zufällige Ausreißer den Bereich der Spannweite leicht verzerren können, werden 5 % der Kontrollfälle automatisch an jedem Ende weggelassen.

Zwei unabhängige Stichproben: Gruppen definieren

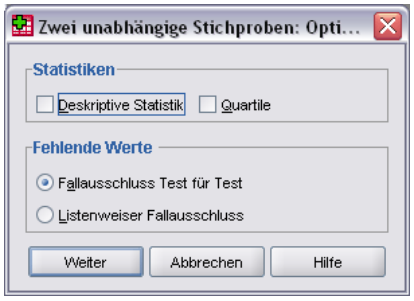
Abbildung 36-10
Dialogfeld "Zwei unabhängige Stichproben: Gruppen definieren"



Um die Datei in zwei Gruppen oder Stichproben aufzuteilen, geben Sie eine ganze Zahl für Gruppe 1 und eine weitere Zahl für Gruppe 2 ein. Fälle mit anderen Werten werden aus der Analyse ausgeschlossen.

Tests bei zwei unabhängigen Stichproben – Optionen

Abbildung 36-11
Dialogfeld "Zwei unabhängige Stichproben: Optionen"



Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Zeigt Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle an.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluss Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (Tests bei zwei unabhängigen Stichproben)

Mit dem Unterbefehl `MOSES` der Befehlssyntax-Sprache kann die Anzahl der Fälle angegeben werden, die für den Moses-Test getrimmt werden sollen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Tests bei zwei verbundenen Stichproben

Die Prozedur "Tests bei zwei verbundenen Stichproben" vergleicht die Verteilungen von zwei Variablen.

Beispiel. Erhalten Familien, die ihr Haus verkaufen, im allgemeinen den geforderten Preis? Wenn Sie den Wilcoxon-Test auf die Daten von 10 Häusern anwenden, könnten Sie beispielsweise feststellen, dass sieben Familien weniger als den geforderten Preis, eine Familie mehr als den geforderten Preis und zwei Familien den geforderten Preis erhielten.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile. Tests: Wilcoxon-Test, Vorzeichentest, McNemar. Wenn die Option "Exakte Tests" installiert ist (nur unter Windows-Betriebssystemen verfügbar) steht außerdem der Rand-Homogenitätstest zur Verfügung.

Daten. Verwenden Sie numerische Variablen, die geordnet werden können.

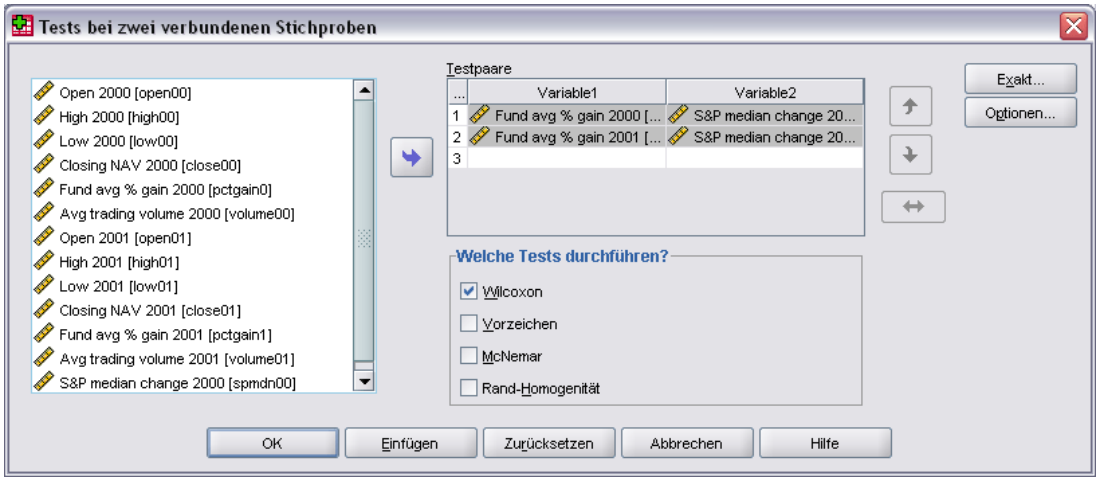
Annahmen. Obwohl keine bestimmten Verteilungen für die beiden Variablen vorausgesetzt werden, wird die Verteilung der Grundgesamtheit der gepaarten Differenzen als symmetrisch angenommen.

So lassen Sie Tests bei zwei verbundenen Stichproben berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    Zwei verbundene Stichproben…

Abbildung 36-12
Dialogfeld "Tests bei zwei verbundenen Stichproben"



▶ Wählen Sie mindestens ein Variablenpaar aus.

Typen von Tests bei zwei verbundenen Stichproben

Die Tests in diesem Abschnitt vergleichen die Verteilungen von zwei verbundenen Variablen. Der geeignete Test hängt vom jeweiligen Datentyp ab.

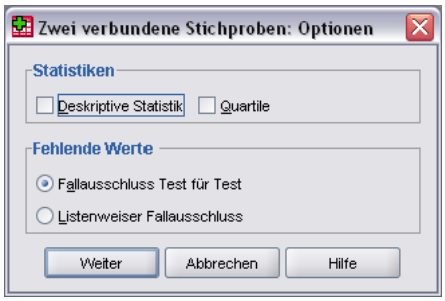
Falls Ihre Daten stetig sind, verwenden Sie den Vorzeichentest oder den Wilcoxon-Test. Der **Vorzeichentest** berechnet für alle Fälle die Differenzen zwischen den beiden Variablen und klassifiziert sie als positiv, negativ oder verbunden. Falls die beiden Variablen ähnlich verteilt sind, unterscheidet sich die Zahl der positiven und negativen Differenzen nicht signifikant. Der **Wilcoxon-Test** berücksichtigt sowohl Informationen über Vorzeichen der Differenzen als auch die Größe der Differenzen zwischen den Paaren. Da der Wilcoxon-Test mehr Informationen über die Daten aufnimmt, kann er mehr leisten als der Vorzeichentest.

Falls Sie mit binären Daten arbeiten, verwenden Sie den **McNemar-Test**. Dieser Test wird üblicherweise bei Messwiederholungen verwendet, wenn jede Antwort eines Subjektes doppelt abgerufen wird, einmal bevor ein festgelegtes Ereignis eintritt und einmal danach. Der McNemar-Test bestimmt, ob die Antwortrate am Anfang (vor dem Ereignis) gleich der Antwortrate am Ende (nach dem Ereignis) ist. Dieser Test ist für das Erkennen von Änderungen bei Antworten nützlich, die durch experimentelle Einflußnahme in sogenannten "Vorher-und-nachher-Designs" entstanden sind.

Falls Sie mit kategorialen Daten arbeiten, verwenden Sie den **Rand-Homogenitätstest**. Dieser Test ist eine Erweiterung des McNemar-Tests von binären Variablen auf multinomiale Variablen. Mithilfe dieses Tests wird unter Verwendung der Chi-Quadrat-Verteilung überprüft, ob Änderungen bei den Antworten vorliegen. Dies ist nützlich, um zu ermitteln, ob die Änderungen in sogenannten "Vorher-und-nachher-Designs" durch experimentelle Einflußnahme verursacht werden. Der Rand-Homogenitätstest ist nur verfügbar, wenn Sie die SPSS Exact Tests installiert haben.

Optionen für Tests bei zwei verbundenen Stichproben

Abbildung 36-13
Dialogfeld "Zwei verbundene Stichproben: Optionen"



Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Zeigt Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle an.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluss Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (zwei verbundene Stichproben)

Mit der Befehlssyntax-Sprache können Sie außerdem eine Variable mit jeder Variable auf einer Liste überprüfen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Tests bei mehreren unabhängigen Stichproben

Mit der Prozedur "Tests bei mehreren unabhängigen Stichproben" werden zwei oder mehrere Fallgruppen einer Variablen verglichen.

Beispiel. Unterscheiden sich 100-Watt-Glühlampen dreier Marken in ihrer durchschnittlichen Lebensdauer? Mit der einfaktoriellen Varianzanalyse nach Kruskal-Wallis könnten Sie feststellen, dass die drei Marken sich in ihrer durchschnittlichen Lebensdauer unterscheiden.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile. Tests: Kruskal-Wallis-*H*, Median.

Daten. Verwenden Sie numerische Variablen, die geordnet werden können.

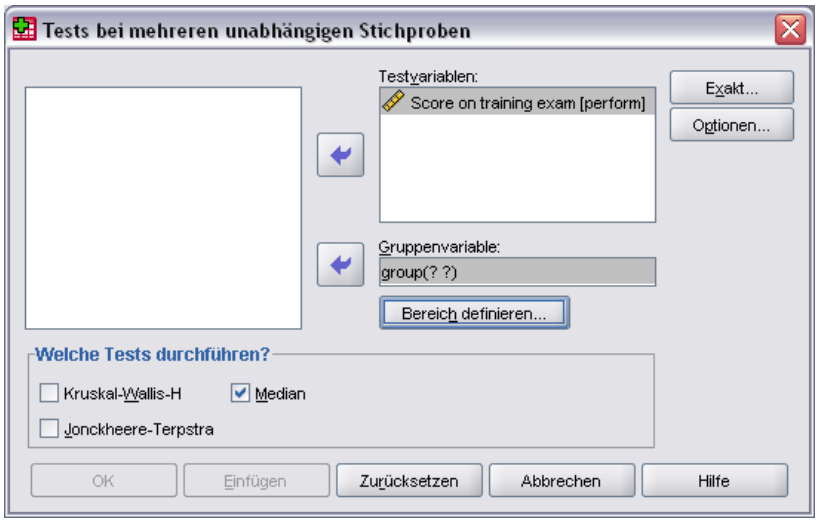
Annahmen. Verwenden Sie unabhängige Zufallsstichproben. Für den Kruskal-Wallis-*H*-Test sind Stichproben erforderlich, die sich in ihrer Form ähneln.

So lassen Sie Tests für mehrere unabhängige Stichproben berechnen:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Nichtparametrische Tests
  K unabhängige Stichproben...

Abbildung 36-14
Festlegung des Mediantests



- Wählen Sie mindestens eine numerische Variable aus.
- Wählen Sie eine Gruppenvariable aus und klicken Sie auf Bereich definieren, um die ganzzahligen Minimal- und Maximalwerte der Gruppenvariablen festzulegen.

Tests bei mehreren unabhängigen Stichproben: Welche Tests durchführen?

Sie können mit drei Tests bestimmen, ob mehrere unabhängige Stichproben aus derselben Grundgesamtheit stammen. Mit dem Kruskal-Wallis-*H*-Test, dem Mediantest und dem Jonckheere-Terpstra-Test können Sie prüfen, ob mehrere unabhängige Stichproben aus derselben Grundgesamtheit stammen.

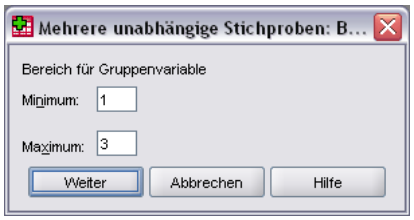
Der **Kruskal-Wallis-H-Test**, eine Erweiterung des Mann-Whitney-*U*-Tests, ist die nichtparametrische Entsprechung der einfaktoriellen Varianzanalyse und erkennt Unterschiede in der Lage der Verteilung. Der **Mediantest**, der allgemeiner, aber nicht so leistungsstark ist, erkennt Unterschiede von Verteilungen in Lage und Form. Der Kruskal-Wallis-*H*-Test und der Mediantest setzen voraus, dass keine *a-priori*-Ordnung der *k* Grundgesamtheiten vorliegt, aus denen die Stichproben gezogen werden.

Wenn eine natürliche *a-priori*-Ordnung (aufsteigend oder absteigend) der *k* Grundgesamtheiten *besteht*, ist der **Jonckheere-Terpstra-Test** leistungsfähiger. Die *k* Grundgesamtheiten könnten zum Beispiel *k* ansteigende Temperaturen darstellen. Die Hypothese, dass unterschiedliche Temperaturen die gleiche Verteilung von Antworten erzeugen, wird gegen die Alternative getestet, dass mit Zunahme der Temperatur die Größe der Antwort zunimmt. Hierbei ist die alternative Hypothese geordnet, deshalb ist der Jonckheere-Terpstra-Test für diesen Test am besten geeignet.

Der Jonckheere-Terpstra-Test ist nur verfügbar, wenn Sie das Erweiterungsmodul Exact Tests installiert haben.

Tests bei mehreren unabhängigen Stichproben: Bereich definieren

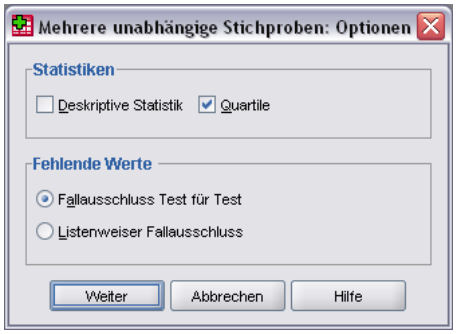
Abbildung 36-15
Dialogfeld "Mehrere unabhängige Stichproben: Bereich definieren"



Um den Bereich zu definieren, geben Sie für Minimum und Maximum ganzzahlige Werte ein, die der niedrigsten und höchsten Kategorie der Gruppenvariablen entsprechen. Der Minimalwert muss kleiner sein als der Maximalwert. Wenn Sie zum Beispiel als Minimum 1 und als Maximum 3 angeben, werden nur die ganzzahligen Werte von 1 bis 3 verwendet. Das Minimum muss kleiner als das Maximum sein. Beide Werte müssen angegeben werden.

Tests bei mehreren unabhängigen Stichproben: Optionen

Abbildung 36-16
Dialogfeld "Mehrere unabhängige Stichproben: Optionen"



Statistik. Sie können eine oder beide Auswertungsstatistiken wählen.

- **Deskriptive Statistik.** Zeigt Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle an.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Fehlende Werte. Bestimmt die Verarbeitung fehlender Werte.

- **Fallausschluss Test für Test.** Werden mehrere Tests festgelegt, so wird jeder Test einzeln auf fehlende Werte geprüft.
- **Listenweiser Fallausschluss.** Fälle mit fehlenden Werten für eine Variable werden aus allen Analysen ausgeschlossen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (K unabhängige Stichproben)

In der Befehlssyntax-Sprache haben Sie außerdem die Möglichkeit, mit dem Unterbefehl MEDIAN einen anderen Wert als den beobachteten Median für den Mediantest festzulegen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Tests bei mehreren verbundenen Stichproben

Bei der Prozedur "Tests bei mehreren verbundenen Stichproben" werden die Verteilungen von zwei oder mehr Variablen verglichen.

Beispiel. Genießen die Berufsgruppen Ärzte, Anwälte, Polizisten oder Lehrer in der Öffentlichkeit ein unterschiedliches Ansehen? Zehn Personen wurden gebeten, diese vier Berufsgruppen in der Reihenfolge ihres Ansehens anzuordnen. Der Test nach Friedman zeigt, dass diese vier Berufsgruppen in der Öffentlichkeit tatsächlich ein unterschiedliches Ansehen genießen.

Statistiken. Mittelwert, Standardabweichung, Minimum, Maximum, Anzahl der nichtfehlenden Fälle und Quartile. Tests: Friedman, Kendall-*W* und Cochran-*Q*.

Daten. Verwenden Sie numerische Variablen, die geordnet werden können.

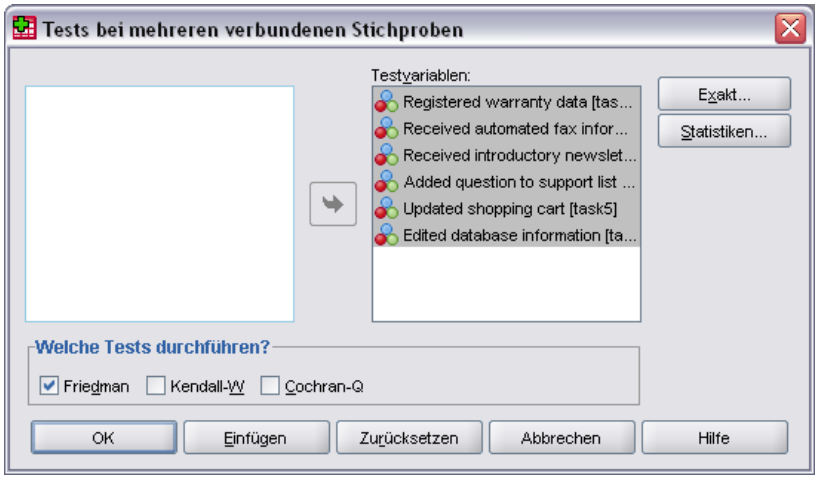
Annahmen. Nichtparametrische Tests erfordern keine Annahmen über die Form der zugrunde liegenden Verteilung. Verwenden Sie abhängige Zufallsstichproben.

So lassen Sie Tests bei mehreren verbundenen Stichproben berechnen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Nichtparametrische Tests
    K verbundene Stichproben...

Abbildung 36-17
Auswahl von "Cochran" als Testtyp



▶ Wählen Sie zwei oder mehr numerische Testvariablen aus.

Tests bei mehreren verbundenen Stichproben: Welche Tests durchführen?

Sie können die Verteilung von verschiedenen verbundenen Variablen mit drei Tests vergleichen.

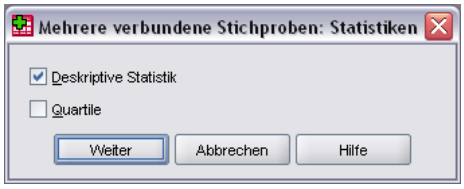
Der **Test** nach **Friedman** stellt das nichtparametrische Äquivalent eines Designs mit Messwiederholungen bei einer Stichprobe bzw. eine Zweifach-Varianzanalyse mit einer Beobachtung pro Zelle dar. Der Friedman-Test überprüft die Nullhypothese, wonach die *k* verbundenen Variablen aus derselben Grundgesamtheit stammen. Für jeden Fall werden den *k* Variablen Rangzahlen von 1 bis *k* zugewiesen. Die Teststatistik wird auf der Grundlage dieser Ränge durchgeführt.

Das **Kendall-W** stellt eine Normalisierung der Statistik nach Friedman dar. Das Kendall-*W* kann als Konkordanzkoeffizient interpretiert werden, der ein Maß für die Übereinstimmung der Prüfer darstellt. Jeder Fall ist ein Richter oder Prüfer, und jede Variable ist ein zu beurteilendes Objekt oder eine zu beurteilende Person. Die Rangsumme jeder Variablen wird berechnet. Das Kendall-*W* liegt im Bereich von 0 (keine Übereinstimmung) bis 1 (vollständige Übereinstimmung).

Das **Cochran-Q** entspricht vollständig dem Friedman-Test. Es wird jedoch angewendet, wenn alle Antworten binär sind. Dieser Test stellt eine Erweiterung des McNemar-Tests auf *k* Stichproben dar. Das Cochran-*Q* überprüft die Hypothese, dass mehrere verbundene dichotome Variablen denselben Mittelwert aufweisen. Die Variablenwerte beziehen sich auf dasselbe Individuum oder auf zusammengehörige Individuen.

Tests bei mehreren verbundenen Stichproben: Statistiken

Abbildung 36-18
Dialogfeld "Mehrere verbundene Stichproben: Statistiken"



Sie können Statistiken auswählen.

- **Deskriptive Statistik.** Zeigt Mittelwert, Standardabweichung, Minimum, Maximum und Anzahl der nichtfehlenden Fälle an.
- **Quartile.** Zeigt die Werte an, die den 25., 50. und 75. Perzentilen entsprechen.

Zusätzliche Funktionen beim Befehl NPAR TESTS (K verbundene Stichproben)

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
37

Analyse von Mehrfachantworten

Sie können für die Analyse von Sets aus dichotomen Variablen und von Sets aus kategorialen Variablen zwei Prozeduren verwenden. Mit der Prozedur "Mehrfachantworten: Häufigkeiten" können Sie Häufigkeitstabellen erstellen. Mit der Prozedur "Mehrfachantworten: Kreuztabellen" werden zwei- oder dreidimensionale Kreuztabellen angezeigt. Sie müssen Mehrfachantworten-Sets definieren, ehe Sie mit einer der Prozeduren beginnen.

Beispiel. Dieses Beispiel veranschaulicht den Gebrauch von Mehrfachantworten in einer Marktforschungsanalyse. Die hier verwendeten Daten sind frei erfunden und dürfen nicht als real interpretiert werden. Eine Fluggesellschaft führt eine Umfrage unter den Passagieren einer bestimmten Flugroute durch, um Informationen über konkurrierende Fluggesellschaften zu erhalten. In diesem Beispiel möchte American Airlines in Erfahrung bringen, welche anderen Fluggesellschaften ihre Passagiere auf der Route Chicago-New York nutzen und welche Rolle der Flugplan sowie der Service bei der Auswahl der Fluggesellschaft spielen. Der Flugbegleiter händigt jedem Passagier beim Einsteigen in die Maschine einen kurzen Fragebogen aus. Die erste Frage lautet: "Kreuzen Sie bitte alle der folgenden Fluggesellschaften an, mit denen Sie diese Route in den letzten sechs Monaten geflogen sind: American, United, TWA, USAir und andere." Dies ist eine Frage, die mit Mehrfachantworten beantwortet werden kann, weil jeder Passagier mehr als eine Antwort ankreuzen kann. Sie können diese Frage aber nicht direkt kodieren, weil eine SPSS-Variable für jeden Fall nur einen Wert annehmen kann. Sie müssen mehrere Variablen verwenden, um die Antworten zu jeder Frage zu erfassen. Dazu haben Sie zwei Möglichkeiten. Eine Möglichkeit besteht darin, zu jeder Antwortmöglichkeit eine entsprechende Variable zu definieren, also zum Beispiel "American", "United", "TWA", "USAir" und "andere". Wenn ein Passagier "United" ankreuzt, wird der Variablen *united* der Code 1 zugewiesen, sonst erhält diese den Code 0. Bei dieser Methode werden Variablen in **mehreren Dichotomien** erfaßt. Eine andere Möglichkeit stellt das Erfassen der Antworten in **mehreren Kategorien** dar, bei der Sie die maximale Anzahl möglicher Antworten auf die Frage schätzen und eine entsprechende Anzahl von Variablen festlegen. Hierbei wird die verwendete Fluggesellschaft mit Hilfe eines Codes angegeben. Beim Durchsehen einer Stichprobe von Fragebögen stellen Sie vielleicht fest, daß in den letzten sechs Monaten kein Passagier mit mehr als drei verschiedenen Fluggesellschaften auf dieser Route geflogen ist. Außerdem bemerken Sie, daß aufgrund der Liberalisierung des Luftverkehrs 10 weitere Fluggesellschaften in der Kategorie "Andere" genannt sind. Mit der Methode für mehrere Kategorien würden Sie drei Variablen definieren. Jede würde wie folgt kodiert sein: 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta* usw. Wenn ein Passagier "American" und "TWA" ankreuzt, wird der ersten Variablen der Code 1 zugewiesen, der zweiten der Code 3 und der dritten ein Code für fehlende Werte. Ein anderer Passagier hat vielleicht "American" und "Delta" angekreuzt. Dementsprechend wird der ersten Variablen der Code 1, der zweiten der Code 5 und der dritten ein Code für fehlende Werte zugewiesen. Dagegen führt die Methode für mehrfache Dichotomie zu 14 verschiedenen Variablen. Obwohl beide Methoden

für dieses Umfragebeispiel geeignet sind, hängt die Wahl der Methode von der Verteilung der Antworten ab.

Mehrfachantworten: Sets definieren

Mit der Prozedur "Mehrfachantworten: Sets definieren" können Sie elementare Variablen in Sets aus dichotomen Variablen und Sets aus kategorialen Variablen gruppieren. Für diese Sets können Sie Häufigkeitstabellen und Kreuztabellen erstellen. Sie können bis zu 20 Mehrfachantworten-Sets definieren. Jedes Set muß über einen eigenen eindeutigen Namen verfügen. Sie können ein Set entfernen, indem Sie es in der Liste der Mehrfachantworten-Sets markieren und anschließend auf Entfernen klicken. Sie können ein Set ändern, indem Sie es in der Liste markieren, die Charakteristiken der Set-Definition ändern und anschließend auf Ändern klicken.

Sie können die elementaren Variablen als Dichotomien oder als Kategorien definieren. Wenn Sie dichotome Variablen verwenden möchten, aktivieren Sie das Optionsfeld Dichotomien, um ein Set aus dichotomen Variablen zu erstellen. Geben Sie für "Gezählter Wert" eine ganze Zahl ein. Jede Variable, bei welcher der gezählte Wert mindestens einmal auftritt, wird zu einer Kategorie des Sets aus dichotomen Variablen. Aktivieren Sie das Optionsfeld Kategorien, um ein Set aus kategorialen Variablen zu erstellen, das den gleichen Wertebereich wie die Komponentenvariablen umfaßt. Geben Sie ganzzahlige Werte für die Minimal- und Maximalwerte des Bereichs für die Kategorien des Sets aus kategorialen Variablen ein. SPSS bildet die Summe aller unterschiedlichen ganzzahligen Werte im Bereich aller Komponentenvariablen. Leere Kategorien werden nicht in Tabellen übernommen.

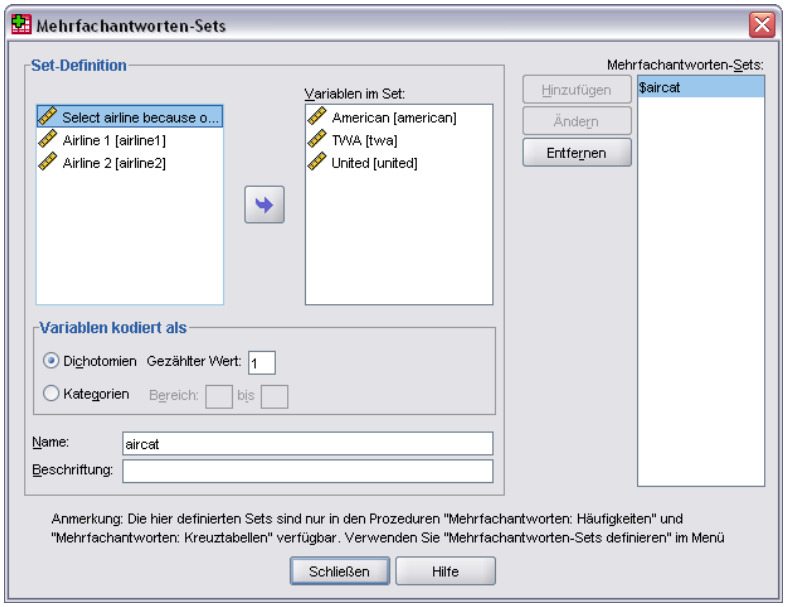
Sie müssen jedem Mehrfachantworten-Set einen eindeutigen Namen zuweisen, der aus bis zu sieben Zeichen bestehen darf. SPSS stellt dem von Ihnen zugewiesenen Namen das Dollarzeichen ($) als Präfix voran. Die folgenden reservierten Namen dürfen Sie nicht verwenden: *casenum*, *sysmis*, *jdate*, *date*, *time*, *length* und *width*. Der Name des Mehrfachantworten-Sets ist nur zur Verwendung in Mehrfachantworten-Prozeduren vorgesehen. In anderen Prozeduren können Sie sich nicht auf Namen von Mehrfachantworten-Sets beziehen. Wahlweise können Sie für das Mehrfachantworten-Set ein aussagekräftiges Variablenlabel eingeben. Das Label kann bis zu 40 Zeichen lang sein.

So definieren Sie Mehrfachantworten-Sets

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mehrfachantworten
    Sets definieren...

Abbildung 37-1
Dialogfeld "Mehrfachantworten-Sets"



- ▶ Wählen Sie mindestens zwei Variablen aus.
- ▶ Wenn Ihre Variablen als Dichotomien kodiert sind, geben Sie an, welcher Wert gezählt werden soll. Wenn Ihre Variablen als Kategorien kodiert sind, legen Sie den Bereich für die Kategorien fest.
- ▶ Geben Sie einen eindeutigen Namen für jedes Mehrfachantworten-Set ein.
- ▶ Klicken Sie auf Hinzufügen, um das Mehrfachantworten-Set zur Liste der definierten Sets hinzuzufügen.

Mehrfachantworten: Häufigkeiten

Mit der Prozedur "Mehrfachantworten: Häufigkeiten" erstellen Sie Häufigkeitstabellen für Mehrfachantworten-Sets. Zuvor müssen Sie mindestens ein Mehrfachantworten-Set definieren (siehe "Mehrfachantworten: Sets definieren").

Bei Sets aus dichotomen Variablen entsprechen die in der Ausgabe gezeigten Kategorienamen den Variablenlabels, die für die elementaren Variablen in der Gruppe festgelegt wurden. Wenn keine Variablenlabels festgelegt wurden, werden die Variablennamen als Labels verwendet. Bei Sets aus kategorialen Variablen entsprechen die Kategoriebeschriftungen den Wertelabels der ersten Variable in der Gruppe. Wenn Kategorien, die bei der ersten Variable fehlen, bei anderen Variablen in der Gruppe vorhanden sind, müssen Sie ein Wertelabel für die fehlenden Kategorien festlegen.

Fehlende Werte. Fälle mit fehlenden Werten werden jeweils für einzelne Tabellen ausgeschlossen. Sie können aber auch eine oder beide der folgenden Möglichkeiten auswählen:

- ■ **Für dichotome Variablen Fälle listenweise ausschließen.** Fälle, bei denen Werte einer beliebigen Variablen fehlen, werden aus der Tabelle des Sets aus dichotomen Variablen ausgeschlossen. Dies gilt nur für Mehrfachantworten-Sets, die als Sets aus dichotomen Variablen definiert

wurden. In der Standardeinstellung gilt ein Fall in einem Set aus dichotomen Variablen als fehlend, wenn keine der Variablen des Falls den gezählten Wert enthält. Fälle mit fehlenden Werten für nur einige, aber nicht alle der Variablen werden in die Tabellen der Gruppe aufgenommen, wenn mindestens eine Variable den gezählten Wert enthält.

- **Für kategoriale Variablen Fälle listenweise ausschließen.** Fälle, bei denen Werte einer beliebigen Variablen fehlen, werden aus der Tabelle des Sets aus kategorialen Variablen ausgeschlossen. Dies gilt nur für Mehrfachantworten-Sets, die als Sets aus kategorialen Variablen definiert wurden. In der Standardeinstellung gilt ein Fall in einem Set aus kategorialen Variablen nur als fehlend, wenn keine der Komponenten des Falls gültige Werte innerhalb des definierten Bereichs enthält.

Beispiel. Jede SPSS-Variable, die aus einer Frage in einer Umfrage erstellt wurde, stellt eine elementare Variable dar. Zum Analysieren der Mehrfachantworten müssen Sie die Variablen in einem der beiden möglichen Typen von Mehrfachantworten-Sets zusammenfassen: in einem Set aus dichotomen Variablen oder in einem Set aus kategorialen Variablen. Wenn zum Beispiel in einer Umfrage ermittelt wurde, mit welcher von drei verschiedenen Fluggesellschaften (American, United und TWA) die befragten Personen in den letzten sechs Monaten geflogen sind, und Sie haben dichotome Variablen verwendet und ein **Set aus dichotomen Variablen** definiert, dann würde jede der drei Variablen im Set zu einer Kategorie der Gruppenvariablen werden. Die Angaben zu Anzahl und Prozentwert für jede Fluggesellschaft werden zusammen in einer Häufigkeitstabelle angezeigt. Wenn Sie feststellen, dass keiner der Befragten mit mehr als zwei Fluggesellschaften geantwortet hat, können Sie zwei Variablen erstellen, die jeweils einen von drei Codes annehmen können. Dabei stellt jeder Code eine Fluggesellschaft dar. Wenn Sie ein **Set aus kategorialen Variablen** definieren, stellen die Werte in der Tabelle die Anzahl von gleichen Codes in den elementaren Variablen dar. Das resultierende Set von Werten entspricht denen für jede einzelne der elementaren Variablen. So entsprechen beispielsweise 30 Antworten mit "United" der Summe von fünf Antworten mit "United" für "Fluglinie 1" und 25 Antworten mit "United" für "Fluglinie 2". Die Angaben zu Anzahl und Prozentwert für jede Fluggesellschaft werden zusammen in einer Häufigkeitstabelle angezeigt.

Statistiken. Häufigkeitstabellen mit den Häufigkeiten, Prozentsätzen der Antworten, Prozentsätzen der Fälle, der Anzahl gültiger Fälle und der Anzahl fehlender Fälle.

Daten. Verwenden Sie Mehrfachantworten-Sets.

Annahmen. Die Häufigkeiten und Prozentsätze geben nützliche Beschreibungen für Daten mit beliebigen Verteilungen.

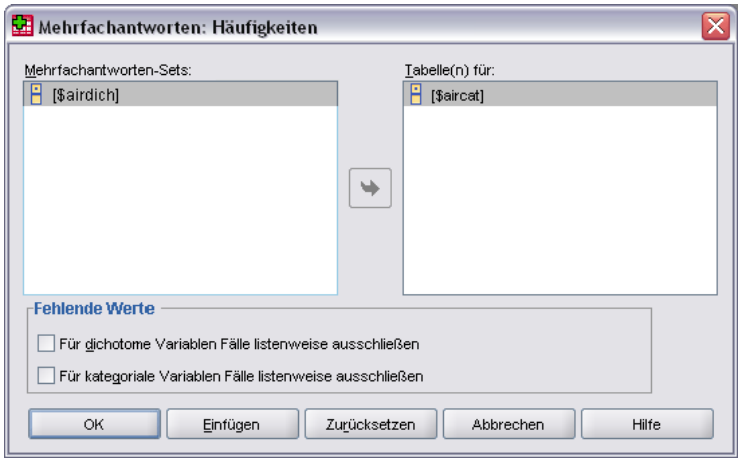
Verwandte Prozeduren. Mit der Prozedur "Mehrfachantworten: Sets definieren" können Sie Mehrfachantworten-Sets definieren.

So berechnen Sie Häufigkeiten mit Mehrfachantworten:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mehrfachantworten
    Häufigkeiten...

Abbildung 37-2
Dialogfeld "Mehrfachantworten: Häufigkeiten"



- Wählen Sie mindestens ein Mehrfachantworten-Set aus.

Mehrfachantworten: Kreuztabellen

Mit der Prozedur "Mehrfachantworten: Kreuztabellen" können Kreuztabellen für definierte Mehrfachantworten-Sets, elementare Variablen oder eine Kombination dieser Elemente berechnet werden. Sie können außerdem Prozentwerte für Zellen basierend auf Fällen oder Antworten berechnen lassen, die Verarbeitung von fehlenden Werten ändern oder gepaarte Kreuztabellen erstellen lassen. Zuvor müssen Sie mindestens ein Mehrfachantworten-Set definieren (siehe "So definieren Sie Mehrfachantworten-Sets").

Bei Sets aus dichotomen Variablen entsprechen die in der Ausgabe gezeigten Kategorienamen den Variablenlabels, die für die elementaren Variablen in der Gruppe festgelegt wurden. Wenn keine Variablenlabels festgelegt wurden, werden die Variablennamen als Labels verwendet. Bei Sets aus kategorialen Variablen entsprechen die Kategoriebeschriftungen den Wertelabels der ersten Variable in der Gruppe. Wenn Kategorien, die bei der ersten Variable fehlen, bei anderen Variablen in der Gruppe vorhanden sind, müssen Sie ein Wertelabel für die fehlenden Kategorien festlegen. In SPSS werden die Kategorienbeschriftungen für Spalten auf drei Zeilen mit bis zu acht Zeichen pro Zeile angezeigt. Wenn Sie vermeiden möchten, dass Wörter getrennt werden, können Sie die Anordnung von Zeilen und Spalten umdrehen oder die Labels neu festlegen.

Beispiel. Sowohl Sets aus dichotomen Variablen als auch Sets aus kategorialen Variablen können bei dieser Prozedur mit anderen Variablen in eine Kreuztabelle eingehen. Bei einer Befragung von Passagieren einer Fluglinie werden die Reisenden um die folgenden Informationen gebeten: Kreuzen Sie bitte alle der folgenden Fluggesellschaften an, mit denen Sie in den letzten sechs Monaten geflogen sind (American, United und TWA). Was ist wichtiger, wenn Sie einen Flug buchen: der Flugplan oder der Service? Wählen Sie nur eine Möglichkeit aus. Nachdem Sie die Daten als Dichotomien oder multiple Kategorien eingegeben und diese in einem Set zusammengefaßt haben, können Sie die Auswahl der Fluggesellschaften zusammen mit der Frage nach Service bzw. Flugplan als Kreuztabelle berechnen lassen.

Statistiken. Kreuztabellen mit Häufigkeiten pro Zelle, Zeile, Spalte und Gesamt sowie Prozentsätzen für Zellen, Zeilen, Spalten und Gesamt. Die Prozentwerte für die Zellen können auf Fällen oder auf Antworten basieren.

Daten. Verwenden Sie Mehrfachantworten-Sets oder numerische kategoriale Variablen.

Annahmen. Die Häufigkeiten und Prozentsätze geben nützliche Beschreibungen für Daten mit beliebigen Verteilungen.

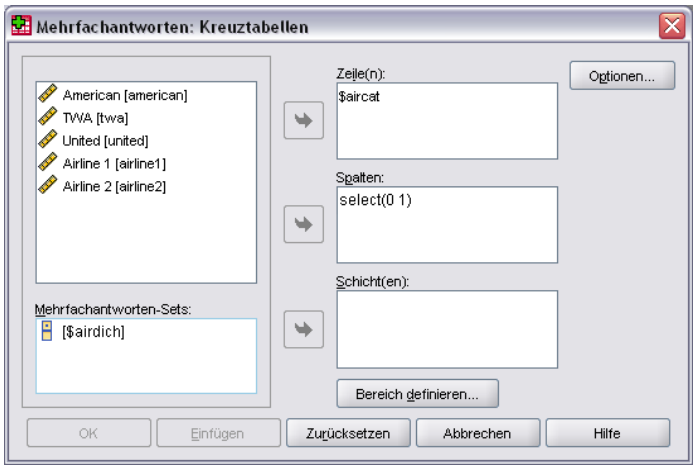
Verwandte Prozeduren. Mit der Prozedur "Mehrfachantworten: Sets definieren" können Sie Mehrfachantworten-Sets definieren.

So berechnen Sie Kreuztabellen mit Mehrfachantworten:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Mehrfachantworten
    Kreuztabellen...

Abbildung 37-3
Dialogfeld "Mehrfachantworten: Kreuztabellen"



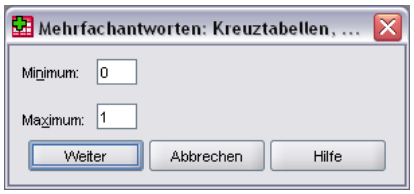
▶ Wählen Sie mindestens eine numerische Variable oder mindestens ein Mehrfachantworten-Set für jede Dimension der Kreuztabelle aus.

▶ Definieren Sie den Bereich jeder elementaren Variablen.

Außerdem können Sie eine Zweifach-Kreuztabelle für jede Kategorie einer Kontroll-Variablen oder eines Mehrfachantworten-Sets berechnen lassen. Wählen Sie mindestens einen Eintrag für die Liste "Schicht(en)" aus.

Mehrfachantworten: Kreuztabellen, Bereich definieren

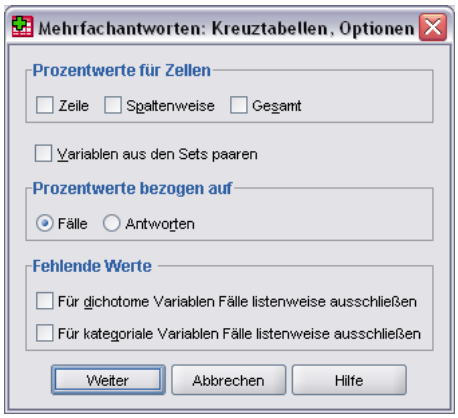
Abbildung 37-4
Dialogfeld "Mehrfachantworten: Kreuztabellen, Bereich definieren"



Für jede elementare Variable in der Kreuztabelle muss ein gültiger Wertebereich festgelegt werden. Geben Sie für die niedrigsten und höchsten Kategoriewerte, die in die Berechnung eingehen sollen, ganze Zahlen ein. Kategorien außerhalb des gültigen Bereichs werden aus der Analyse ausgeschlossen. Bei Werten innerhalb des einschließenden Bereichs wird von ganzen Zahlen ausgegangen, Stellen nach dem Komma werden abgeschnitten.

Mehrfachantworten: Kreuztabellen, Optionen

Abbildung 37-5
Dialogfeld "Mehrfachantworten: Kreuztabellen, Optionen"



Prozentwerte für Zellen. Die Zellenhäufigkeiten werden immer angezeigt. Sie können aber auch Spalten- und Zeilenprozentwerte sowie Prozentwerte für Zweifach-Tabellen (Gesamtwerte) anzeigen lassen.

Prozentwerte bezogen auf. Sie können festlegen, dass die Prozentsätze für die Zellen auf Fällen basieren. Diese Option ist nicht verfügbar, wenn Sie Variablen aus verschiedenen Sets von kategorialen Variablen paaren. Die Prozentsätze für die Zellen können außerdem auf den Antworten basieren. Bei Sets aus dichotomen Variablen entspricht die Anzahl der Antworten der Anzahl von gezählten Werten in allen Fällen. Bei Sets aus kategorialen Variablen entspricht die Anzahl der Antworten der Anzahl von Werten im festgelegten Bereich.

Fehlende Werte. Sie können eine oder beide der folgenden Möglichkeiten auswählen:

- **Für dichotome Variablen Fälle listenweise ausschließen.** Fälle, bei denen Werte einer beliebigen Variablen fehlen, werden aus der Tabelle des Sets aus dichotomen Variablen ausgeschlossen. Dies gilt nur für Mehrfachantworten-Sets, die als Sets aus dichotomen Variablen definiert wurden. In der Standardeinstellung gilt ein Fall in einem Set aus dichotomen Variablen als

fehlend, wenn keine der Variablen des Falls den gezählten Wert enthält. Fälle mit fehlenden Werten für nur einige, aber nicht alle der Variablen werden in die Tabellen der Gruppe aufgenommen, wenn mindestens eine Variable den gezählten Wert enthält.

- **Für kategoriale Variablen Fälle listenweise ausschließen.** Fälle, bei denen Werte einer beliebigen Variablen fehlen, werden aus der Tabelle des Sets aus kategorialen Variablen ausgeschlossen. Dies gilt nur für Mehrfachantworten-Sets, die als Sets aus kategorialen Variablen definiert wurden. In der Standardeinstellung gilt ein Fall in einem Set aus kategorialen Variablen nur als fehlend, wenn keine der Komponenten des Falls gültige Werte innerhalb des definierten Bereichs enthält.

Die Standardeinstellung von SPSS sieht vor, dass beim Erstellen von Kreuztabellen für Sets aus kategorialen Variablen jede Variable in der ersten Gruppe mit jeder Variablen in der zweiten Gruppe gepaart wird und die Häufigkeiten für jede Zelle addiert werden. Deshalb können manche Antworten mehr als einmal in einer Tabelle vorkommen. Sie können die folgende Option auswählen:

Variablen aus den Sets paaren. Hiermit wird die erste Variable aus der ersten Gruppe mit der ersten Variable aus der zweiten Gruppe gepaart usw. Wenn Sie diese Option auswählen, basieren die relativen Häufigkeiten in den Zellen nicht auf den Fällen, sondern auf den Antworten. Bei Sets aus dichotomen Variablen und elementaren Variablen steht das Paaren nicht zur Verfügung.

Zusätzliche Funktionen beim Befehl MULT RESPONSE

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Mit dem Unterbefehl `BY` können Kreuztabellen mit bis zu fünf Dimensionen berechnet werden.
- Mit dem Unterbefehl `FORMAT` können die Optionen für die Ausgabeformatierung geändert werden. So können beispielsweise Wertelabels unterdrückt werden.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
38

Ergebnisberichte

Auflistungen von Fällen und deskriptive Statistiken sind wichtige Hilfsmittel zur Untersuchung und Darstellung von Daten. Mit dem Daten-Editor oder der Prozedur "Berichte" können Sie Fälle auflisten, mit den Prozeduren "Häufigkeiten" Häufigkeitszählungen und deskriptive Statistiken erstellen und mit der Prozedur "Mittelwert" Statistiken für Teilgrundgesamtheiten anfordern. In jeder dieser Prozeduren wird ein zur übersichtlichen Darstellung von Informationen geeignetes Format verwendet. Mit den Funktionen "Bericht in Zeilen" und "Bericht in Spalten" können Sie für Informationen auch ein anderes Format der Datendarstellung wählen.

Bericht in Zeilen

Mit der Funktion "Bericht in Zeilen" werden Berichte erstellt, in denen verschiedene Auswertungsstatistiken in Zeilen angegeben sind. Ebenso sind Listen von Fällen mit oder ohne Auswertungsstatistik verfügbar.

Beispiel. In einem Einzelhandelsunternehmen mit Filialen werden Informationen über Angestellte, Gehälter, Anstellungszeiten sowie Filiale und Abteilung jedes Beschäftigten in Datensätzen gespeichert. Sie können einen Bericht erstellen, der nach Filiale und Abteilung (Break-Variablen) aufgeteilte Informationen (Listen) zu den einzelnen Beschäftigten liefert und eine Auswertungsstatistik (zum Beispiel Durchschnittsgehalt) für jede Filiale, jedes Ressort und jede Abteilung einer Filiale enthält.

Datenspalten. Hier werden die Berichtsvariablen aufgelistet, für die Sie Fälle auflisten oder Auswertungsstatistiken erstellen möchten, und das Anzeigeformat der Datenspalten festgelegt.

Break-Spalten. Hier werden optionale Break-Variablen aufgelistet, die den Bericht in Gruppen aufteilen, und Einstellungen für die Auswertungsstatistik sowie Anzeigeformate für Break-Spalten festgelegt. Bei mehreren Break-Variablen wird für jede Kategorie einer Break-Variablen eine getrennte Gruppe innerhalb der Kategorien der vorhergehenden Break-Variablen in der Liste erzeugt. Die Break-Variablen müssen diskrete kategoriale Variablen sein, welche die Fälle in eine begrenzte Anzahl von sinnvollen Kategorien aufteilen. Die Einzelwerte jeder Break-Variablen werden in einer getrennten Spalte links von allen Datenspalten angezeigt.

Bericht. Hiermit werden alle Merkmale eines Berichts festgelegt, einschließlich zusammenfassender Gesamtstatistiken, Anzeige der fehlenden Werte, Seitennumerierung und Titel.

Fälle anzeigen. Hiermit werden für jeden Fall die aktuellen Werte (oder Wertelabels) von den Variablen der Datenspalten angezeigt. Dadurch wird ein Listenbericht erzeugt, der wesentlich umfangreicher als ein Zusammenfassungsbericht sein kann.

Vorschau. Es wird nur die erste Seite des Berichtes angezeigt. Mit dieser Option erhalten Sie eine Vorschau auf das Format Ihres Berichts, ohne diesen komplett bearbeiten zu müssen.

Daten sind schon sortiert. Bei Berichten mit Break-Variablen muss die Datendatei vor dem Erstellen des Berichts nach den Werten der Break-Variablen sortiert werden. Wenn Ihre Datendatei bereits nach den Werten der Break-Variablen sortiert ist, können Sie durch Auswahl dieser Option Bearbeitungszeit einsparen. Diese Option ist besonders hilfreich, wenn Sie bereits einen Bericht für die Vorschau erstellt haben.

So erstellen Sie eine Zusammenfassung: Bericht in Zeilen

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Berichte
  Bericht in Zeilen...

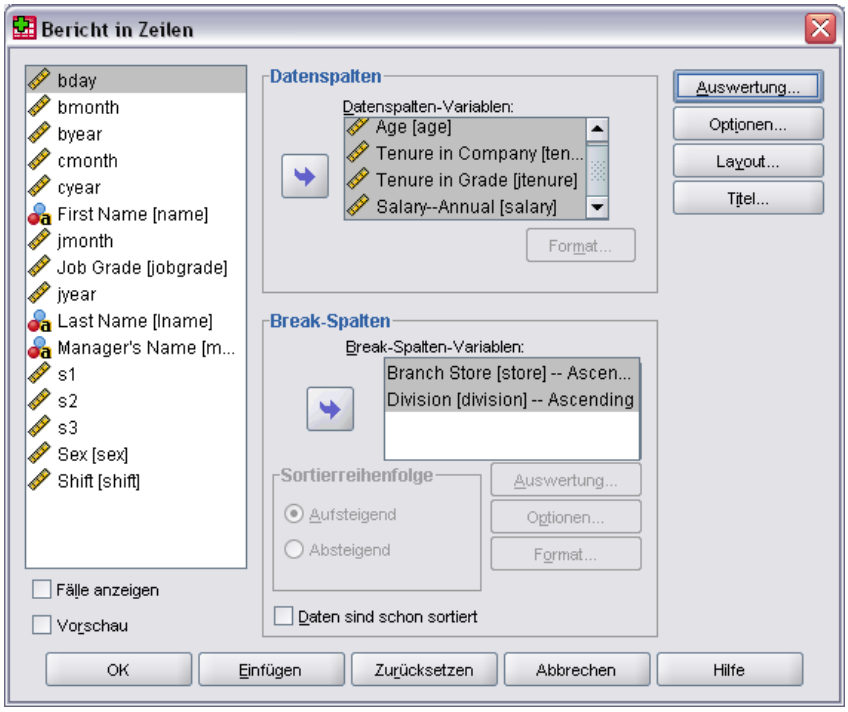
- Wählen Sie mindestens eine Variable für die Datenspalten aus. Für jede ausgewählte Variable wird eine Spalte im Bericht erzeugt.

- Wählen Sie bei sortierten und nach Untergruppen angezeigten Berichten mindestens eine Variable für die Break-Spalten aus.

- Bei Berichten mit Auswertungsstatistiken für Untergruppen, die durch Break-Variablen definiert wurden, wählen Sie in der Liste “Break-Spalten-Variablen” die Break-Variablen aus und klicken Sie im Gruppenfeld “Break-Spalten” auf Auswertung, um das (die) Auswertungsmaß(e) festzulegen.

- Bei Berichten mit zusammenfassenden Auswertungsstatistiken klicken Sie auf Auswertung, um das (die) Auswertungsmaß(e) festzulegen.

Abbildung 38-1
Dialogfeld "Bericht in Zeilen"



Datenspaltenformat/Break-Format in Berichten

In den Format-Dialogfeldern werden Spaltentitel, Spaltenbreite, Textausrichtung sowie Anzeige der Datenwerte oder Wertelabels festgelegt. Mit "Datenspaltenformat" wird das Format der Datenspalten auf der rechten Seite des Berichtes festgelegt. Das Format der Break-Spalten auf der linken Seite wird mit "Break-Format" festgelegt.

Abbildung 38-2
Dialogfeld "Bericht: Datenspaltenformat"

Spaltentitel. Hiermit legen Sie den Spaltentitel für die ausgewählte Variable fest. Lange Titel werden in der Spalte automatisch umgebrochen. Verwenden Sie die Eingabetaste, um Zeilenumbrüche für Titel manuell einzufügen.

Position des Werts in der Spalte. Hiermit wird für die ausgewählte Variable die Ausrichtung des Datenwerts oder Wertelabels in der Spalte festgelegt. Die Ausrichtung der Werte oder Labels hat keinen Einfluß auf die Ausrichtung der Spaltenüberschriften. Der Spalteninhalt kann entweder um eine festgelegte Anzahl von Zeichen eingerückt oder zentriert werden.

Spalteninhalt. Steuert die Anzeige von Datenwerten oder definierten Wertelabels der ausgewählten Variablen. Für Werte ohne definierte Wertelabels werden immer Datenwerte angezeigt. (Nicht verfügbar für Datenspalten in Bericht in Spalten.)

Bericht: Auswertungszeilen für/Endgültige Auswertungszeilen

Die beiden Dialogfelder für Auswertungszeilen legen Einstellungen für die Anzeige der Auswertungsstatistik für Break-Gruppen und für den gesamten Bericht fest. Mit "Auswertung" können Sie Einstellungen bezüglich der Untergruppenstatistik für jede durch die Break-Variablen definierte Kategorie vornehmen. Mit "Endgültige Auswertungszeilen" können Sie Einstellungen für die am Ende des Berichts angezeigte Gesamtstatistik vornehmen.

Abbildung 38-3
Dialogfeld "Bericht: Auswertung"

Die verfügbaren Auswertungsstatistiken sind Summe, Mittelwert, Minimum, Maximum, Anzahl der Fälle, Prozent der Fälle über oder unter einem festgelegten Wert, Prozent der Fälle innerhalb eines festgelegten Wertebereichs, Standardabweichung, Kurtosis, Varianz und Schiefe.

Bericht: Break-Optionen

Mit "Break-Optionen" werden Abstand und Seitenaufteilung der Informationen in den Break-Kategorien festgelegt.

Abbildung 38-4
Dialogfeld "Bericht: Break-Optionen"

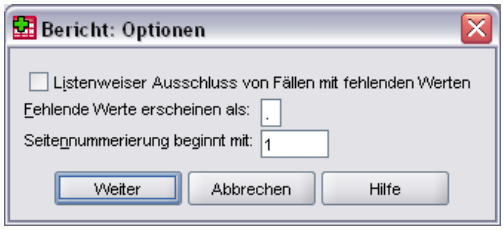
Seiteneinstellung. Hiermit werden Abstand und Seitenaufteilung für Kategorien der ausgewählten Break-Variablen festgelegt. Sie können eine Anzahl von Leerzeilen zwischen den Break-Kategorien festlegen oder eine Break-Kategorie an einen neuen Seitenanfang legen.

Leerzeilen vor Zusammenfassung. Hiermit legen Sie die Anzahl der Leerzeilen zwischen Beschriftungen oder Daten von Break-Kategorien und Auswertungsstatistiken fest. Dies bietet sich besonders für kombinierte Berichte mit Listen von einzelnen Fällen und Auswertungsstatistiken für Break-Kategorien an. In diesen Berichten können Sie Leerraum zwischen Listen von Fällen und Auswertungsstatistiken einfügen.

Bericht: Optionen

Mit "Bericht: Optionen" werden Behandlung und Anzeige der fehlenden Werte sowie Seitenaufteilung des Berichts festgelegt.

Abbildung 38-5
Dialogfeld "Bericht: Optionen"



Fälle mit fehlenden Werten listenweise ausschließen. Für jede der Berichtsvariablen werden sämtliche Fälle mit fehlenden Werten (im Bericht) ausgeschlossen.

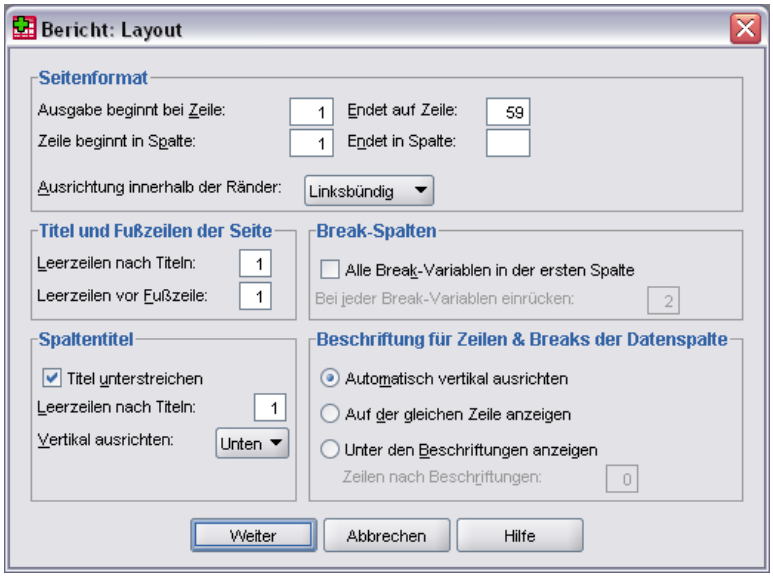
Fehlende Werte erscheinen als. Hier legen Sie das Symbol für fehlende Werte in der Datendatei fest. Das Symbol darf nur aus einem Zeichen bestehen und wird sowohl zur Darstellung **systembedingt fehlender** als auch **benutzerdefiniert fehlender** Werte verwendet.

Seitennumerierung beginnen mit. Mit dieser Option können Sie für die erste Seite des Berichts eine Seitennummer festlegen.

Bericht: Layout

Mit "Bericht: Layout" werden Breite und Länge jeder Berichtsseite, Seitenanordnung des Berichts sowie Einfügen von Leerzeilen und Beschriftungen festgelegt.

Abbildung 38-6
Dialogfeld "Bericht: Layout"



Seitenformat. Legt die Seitenränder, ausgedrückt in Zeilen (oben und unten) und Leerzeichen (links und rechts) sowie die Ausrichtung der Berichte innerhalb der Ränder fest.

Titel und Fußzeilen der Seite. Legt die Anzahl von Zeilen fest, welche die Kopf- und Fußzeile jeweils vom Text des Berichts trennen.

Break-Spalten. Hiermit wird die Anzeige der Break-Spalten festgelegt. Wenn mehrere Break-Variablen festgelegt wurden, können sie sich in getrennten Spalten oder in der ersten Spalte befinden. Das Anordnen aller Break-Variablen in der ersten Spalte erzeugt einen schmaleren Bericht.

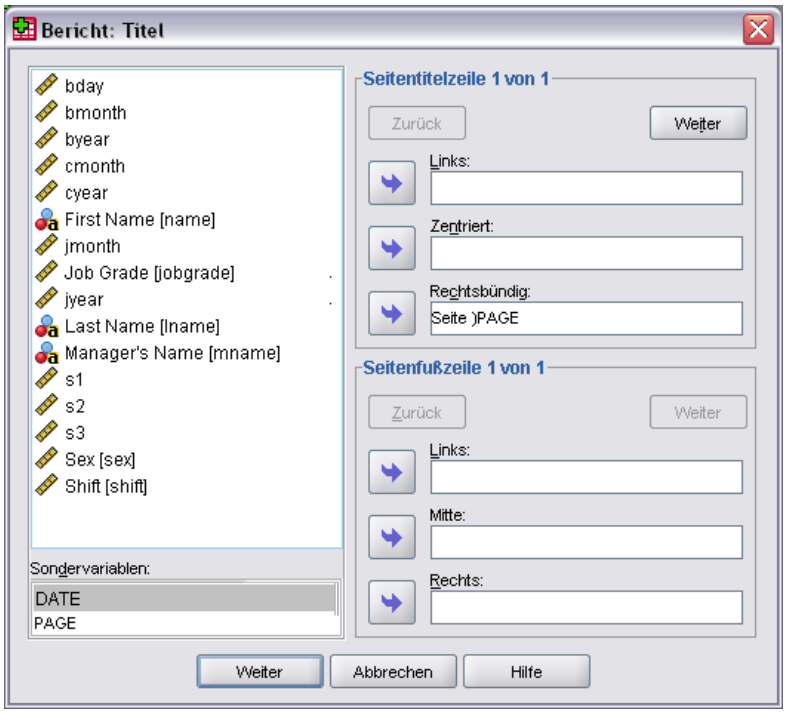
Spaltentitel. Legt die Anzeige von Spaltentiteln fest und umfaßt Unterstreichung des Titels, Anzahl von Leerzeilen zwischen Titel und Text des Berichts sowie die vertikale Ausrichtung.

Beschriftung für Zeilen und Breaks der Datenspalte. Steuert die Anordnung von Informationen in Datenspalten (Datenwerte und/oder Auswertungsstatistiken) bezüglich der Break-Beschriftungen zu Beginn jeder Break-Kategorie. Die erste Informationszeile in der Datenspalte kann entweder in der gleichen Zeile wie die Beschriftung der Break-Kategorie oder nach einer festgelegten Anzahl von Zeilen nach der Beschriftung der Break-Kategorie beginnen. (Nicht für Auswertungsberichte in Spalten verfügbar.)

Bericht: Titel

Im Dialogfeld "Bericht: Titel" werden Inhalt und Anordnung der Titel- und Fußzeilen des Berichts festgelegt. Sie können jeweils bis zu zehn Titel- und Fußzeilen festlegen, wobei in jeder Zeile linksbündige, zentrierte oder rechtsbündige Komponenten enthalten sein können.

Abbildung 38-7
Dialogfeld "Bericht: Titel"



Wenn Sie in Titeln oder Fußzeilen Variablen eingeben, wird das aktuelle Wertelabel oder der Wert der Variablen im Titel oder in der Fußzeile angezeigt. In Titeln wird das Wertelabel angezeigt, das dem Wert der Variablen am Beginn der Seite entspricht. In den Fußzeilen wird das Wertelabel angezeigt, das dem Wert der Variablen am Ende der Seite entspricht. Ist kein Wertelabel vorhanden, wird der aktuelle Wert angezeigt.

Sondervariablen. Mit den Sondervariablen *DATE* und *PAGE* können Sie das aktuelle Datum oder die Seitenzahl in eine beliebige Zeile des Kopf- oder Fußzeilenbereichs des Berichts eingeben. Wenn Ihre Datendatei Variablen wie *DATE* oder *PAGE* enthält, können Sie diese in Titeln oder Fußzeilen des Berichts nicht verwenden.

Bericht in Spalten

Mit "Bericht in Spalten" werden Auswertungsberichte erstellt, die in verschiedenen Spalten unterschiedliche Auswertungsstatistiken enthalten.

Beispiel. In einem Einzelhandelsunternehmen mit Filialen werden Informationen über Angestellte, Gehälter, Anstellungszeiten sowie Filiale und Abteilung jedes Beschäftigten in Datensätzen gespeichert. Sie können einen Bericht erstellen, der eine zusammenfassende Gehaltsstatistik (zum Beispiel Mittelwert, Minimum und Maximum) für jede Abteilung liefert.

Datenspalten. Hier werden die Berichtsvariablen aufgelistet, für die Sie eine Auswertungsstatistik anfordern möchten, und das Anzeigeformat sowie die für jede Variable angezeigte Auswertungsstatistik festgelegt.

Break-Spalten. Hiermit werden optionale Break-Variablen, die den Bericht in Gruppen aufteilen, aufgelistet und das Anzeigeformat der Break-Spalten festgelegt. Bei mehreren Break-Variablen wird für jede Kategorie einer Break-Variablen eine getrennte Gruppe innerhalb der Kategorien der vorhergehenden Break-Variablen in der Liste erzeugt. Die Break-Variablen müssen diskrete kategoriale Variablen sein, welche die Fälle in eine begrenzte Anzahl von sinnvollen Kategorien aufteilen.

Bericht. Hiermit legen Sie alle Merkmale des Berichts fest, beispielsweise die Anzeige der fehlenden Werte, Seitennumerierung und Titel.

Vorschau. Es wird nur die erste Seite des Berichtes angezeigt. Mit dieser Option erhalten Sie eine Vorschau auf das Format Ihres Berichts, ohne diesen komplett bearbeiten zu müssen.

Daten sind schon sortiert. Bei Berichten mit Break-Variablen muss die Datendatei vor dem Erstellen des Berichts nach den Werten der Break-Variablen sortiert werden. Wenn Ihre Datendatei bereits nach den Werten der Break-Variablen sortiert ist, können Sie durch Auswahl dieser Option Bearbeitungszeit einsparen. Diese Option ist besonders hilfreich, wenn Sie bereits einen Bericht für die Vorschau erstellt haben.

So erstellen Sie eine Zusammenfassung: Bericht in Spalten

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Berichte
  Bericht in Spalten...

- Wählen Sie mindestens eine Variable für die Datenspalten aus. Für jede ausgewählte Variable wird eine Spalte im Bericht erzeugt.

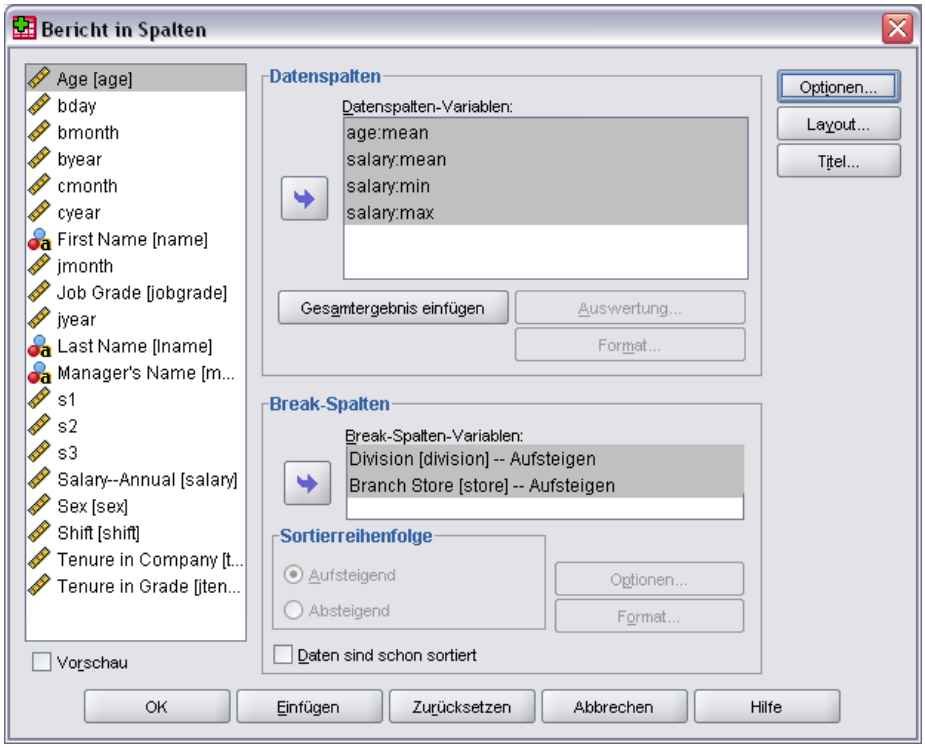
- Um das Auswertungsmaß für eine Variable zu ändern, wählen Sie die Variable in der Liste “Datenspalten-Variablen” aus und klicken Sie auf Auswertung.

- Um mehr als ein Auswertungsmaß für eine Variable berechnen zu lassen, wählen Sie die Variable in der Quellliste aus und übernehmen diese für jedes gewünschte Auswertungsmaß in die Liste “Datenspalten-Variablen”.

- Um eine Spalte mit Summe, Mittelwert, Verhältnis oder einer anderen Funktion einer vorhandenen Spalte anzuzeigen, klicken Sie auf Gesamtergebnis einfügen. Dadurch wird die Variable *Gesamt* in die Liste “Datenspalten” aufgenommen.

- Wählen Sie bei sortierten und nach Untergruppen angezeigten Berichten mindestens eine Variable für die Break-Spalten aus.

Abbildung 38-8
Dialogfeld "Bericht in Spalten"



Datenspalten: Auswertungsfunktion

Im Dialogfeld "Auswertung" wird die angezeigte Auswertungsstatistik der ausgewählten Datenspalten-Variablen festgelegt.

Abbildung 38-9
Dialogfeld "Bericht: Auswertung"

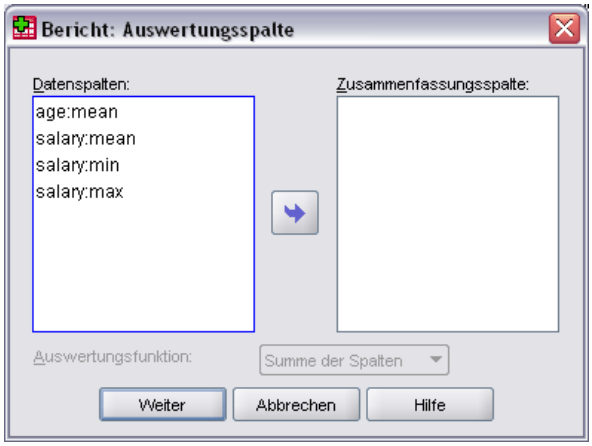
Die verfügbaren Auswertungsstatistiken sind Summe, Mittelwert, Minimum, Maximum, Anzahl der Fälle, Prozent der Fälle über oder unter einem festgelegten Wert, Prozent der Fälle innerhalb eines festgelegten Wertebereichs, Standardabweichung, Varianz, Kurtosis und Schiefe.

Auswertungsspalte für Gesamtergebnis

Im Dialogfeld "Bericht: Auswertungsspalte" werden Einstellungen für die Gesamt-Auswertungsstatistik festgelegt, die zwei oder mehr Datenspalten zusammenfaßt.

Die folgenden Gesamt-Auswertungsstatistiken sind verfügbar: Summe der Spalten, Mittelwert der Spalten, Minimum, Maximum, Differenz zwischen den Werten zweier Spalten, Quotient der Werte in einer Spalte dividiert durch die Werte einer anderen Spalte und das Produkt der miteinander multiplizierten Spaltenwerte.

Abbildung 38-10
Dialogfeld "Bericht: Auswertungsspalte"



Summe der Spalten. Die Spalte *Gesamt* enthält die Summe der Spalten in der Liste "Zusammenfassungsspalte".

Mittelwert der Spalten. Die Spalte *Gesamt* enthält den Durchschnitt der Spalten in der Liste "Zusammenfassungsspalte".

Minimum der Spalten. Die Spalte *Gesamt* enthält den Minimalwert der Spalten in der Liste "Zusammenfassungsspalte".

Maximum der Spalten. Die Spalte *Gesamt* enthält den Maximalwert der Spalten in der Liste "Zusammenfassungsspalte".

1. Spalte – 2. Spalte. Die Spalte *Gesamt* enthält die Differenz zwischen den Spalten in der Liste "Zusammenfassungsspalte". Die Liste "Zusammenfassungsspalte" muss dabei genau zwei Spalten enthalten.

1. Spalte / 2. Spalte. Die Spalte *Gesamt* enthält den Quotienten der Spalten in der Liste "Zusammenfassungsspalte". Die Liste "Zusammenfassungsspalte" muss dabei genau zwei Spalten enthalten.

% 1. Spalte / 2. Spalte. Die Spalte *Gesamt* enthält den prozentualen Anteil der ersten Spalte an der zweiten Spalte in der Liste "Zusammenfassungsspalte". Die Liste "Zusammenfassungsspalte" muss dabei genau zwei Spalten enthalten.

Produkt der Spalten. Die Spalte *Gesamt* enthält das Produkt der Spalten in der Liste "Zusammenfassungsspalte".

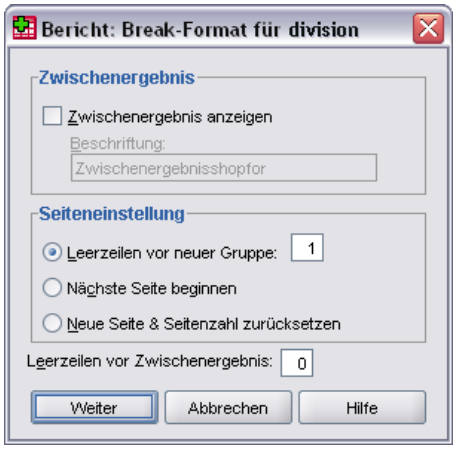
Format der Berichtsspalte

Die Formatoptionen von Daten- und Break-Spalten für "Bericht in Spalten" entsprechen den Optionen für "Bericht in Zeilen".

Bericht: Break-Optionen für Bericht in Spalten

Mit "Break-Optionen" werden Anzeige der Zwischenergebnisse, Abstand und Seitennumerierung für Break-Kategorien festgelegt.

Abbildung 38-11
Dialogfeld "Bericht: Break-Optionen"



Zwischenergebnis. Hiermit wird die Anzeige der Zwischenergebnisse für Break-Kategorien festgelegt.

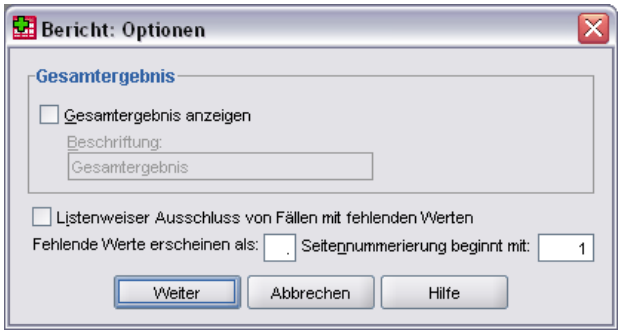
Seiteneinstellung. Hiermit werden Abstand und Seitenaufteilung für Kategorien der ausgewählten Break-Variablen festgelegt. Sie können eine Anzahl von Leerzeilen zwischen den Break-Kategorien festlegen oder eine Break-Kategorie an einen neuen Seitenanfang legen.

Leerzeilen vor Zwischenergebnis. Hiermit legen Sie die Anzahl leerer Zeilen zwischen den Daten der Break-Kategorien und den Zwischenergebnissen fest.

Bericht: Optionen für Bericht in Spalten

Mit "Optionen" werden Anzeige der Gesamtergebnisse, Anzeige der fehlenden Werte und Seitennumerierung in Auswertungsberichten in Spalten festgelegt.

Abbildung 38-12
Dialogfeld "Bericht: Optionen"



Gesamtergebnis. In jeder Spalte wird am unteren Rand ein Gesamtergebnis angezeigt und beschriftet.

Fehlende Werte. Sie können fehlende Werte vom Bericht ausschließen oder fehlende Werte mit einem ausgewählten Zeichen im Bericht kennzeichnen.

Bericht: Layout für Bericht in Spalten

Die Layout-Optionen für "Bericht in Spalten" entsprechen den Optionen für "Bericht in Zeilen".

Zusätzliche Funktionen beim Befehl REPORT

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- In den Spalten einer einzelnen Auswertungszeile lassen sich unterschiedliche Auswertungsfunktionen anzeigen.
- In Datenspalten können Auswertungszeilen für Variablen eingefügt werden, die nicht den Variablen der Datenspalten entsprechen. Außerdem können Zeilen für verschiedene Kombinationen (zusammengesetzte Funktionen) der Auswertungsfunktion eingefügt werden.
- Als Auswertungsfunktionen können Median, Modalwert, Häufigkeit und Prozent verwendet werden.
- Das Anzeigeformat der Auswertungsstatistiken kann genauer festgelegt werden.
- An verschiedenen Stellen des Berichtes können Leerzeilen eingefügt werden.
- In Listenberichten können nach jedem *n*-ten Fall Leerzeilen eingefügt werden.

Wegen der Komplexität der Syntax zum Befehl `REPORT` kann es hilfreich sein, beim Erstellen eines neuen Berichts mit Syntax auf einen vorhandenen Bericht zurückzugreifen. Zum Anpassen eines aus Dialogfeldern erstellten Berichts kopieren Sie die entsprechende Syntax, fügen diese ein und ändern sie so, dass Sie den gewünschten Bericht erstellen können.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
39

Reliabilitätsanalyse

Die Reliabilitätsanalyse ermöglicht es Ihnen, die Eigenschaften von Messniveaus und der Items zu untersuchen, aus denen diese sich zusammensetzen. Mit der Prozedur "Reliabilitätsanalyse" können Sie eine Anzahl von allgemein verwendeten Reliabilitäten des Messniveaus berechnen, und es werden Ihnen Informationen über die Beziehungen zwischen den Items in der Skala zur Verfügung gestellt. Korrelationskoeffizienten in Klassen können verwendet werden, um Reliabilitätsschätzer der Urteiler zu berechnen.

Beispiel. Wird die Kundenzufriedenheit mit Ihrem Fragebogen sinnvoll gemessen? Mit der Reliabilitätsanalyse können Sie das Ausmaß des Zusammenhangs zwischen den Items in Ihrem Fragebogen bestimmen, einen globalen Index der Reproduzierbarkeit bzw. der inneren Konsistenz der vollständigen Skala ermitteln und die kritischen Items herausfinden, welche nicht mehr in der Skala verwendet werden sollten.

Statistiken. Deskriptive Statistiken für jede Variable und für die Skala, Auswertungsstatistik für mehrere Items, Inter-Item-Korrelationen und Inter-Item-Kovarianzen, Reliabilitätsschätzer, ANOVA-Tabelle, Korrelationskoeffizient in Klassen, T^2 nach Hotelling und Tukey-Additivitätstest.

Modelle. Die folgenden Reliabilitätsmodelle sind verfügbar:

- **Alpha (Cronbach).** Dieses Modell ist ein Modell der inneren Konsistenz, welches auf der durchschnittlichen Inter-Item-Korrelation beruht.
- **Split-Half.** Bei diesem Modell wird die Skala in zwei Hälften geteilt und die Korrelation zwischen den Hälften berechnet.
- **Guttman.** Bei diesem Modell werden Guttmans untere Grenzen für die wahre Reliabilität berechnet.
- **Parallel.** Bei diesem Modell wird angenommen, dass alle Items gleiche Varianzen und gleiche Fehlervarianzen für mehrere Wiederholungen aufweisen.
- **Streng parallel.** Bei diesem Modell gelten die Annahmen des parallelen Modells, und es wird zusätzlich die Gleichheit der Mittelwerte der Items angenommen.

Daten. Die Daten können dichotom, ordinal- oder intervallskaliert sein. Sie müssen jedoch numerisch kodiert sein.

Annahmen. Die Beobachtungen sollten unabhängig sein, und Fehler dürfen zwischen den Items nicht korrelieren. Jedes Paar von Items sollte bivariat normalverteilt sein. Die Skalen sollten additiv sein, sodass sich jedes Item linear zum Gesamtwert verhält.

Verwandte Prozeduren. Wenn Sie die Dimensionalität der Skalen-Items untersuchen möchten (um herauszufinden, ob mehr als eine Konstruktion nötig ist, um das Muster der Item-Werte zu erklären), verwenden Sie die Prozedur "Faktorenanalyse" oder "Multidimensionale Skalierung".

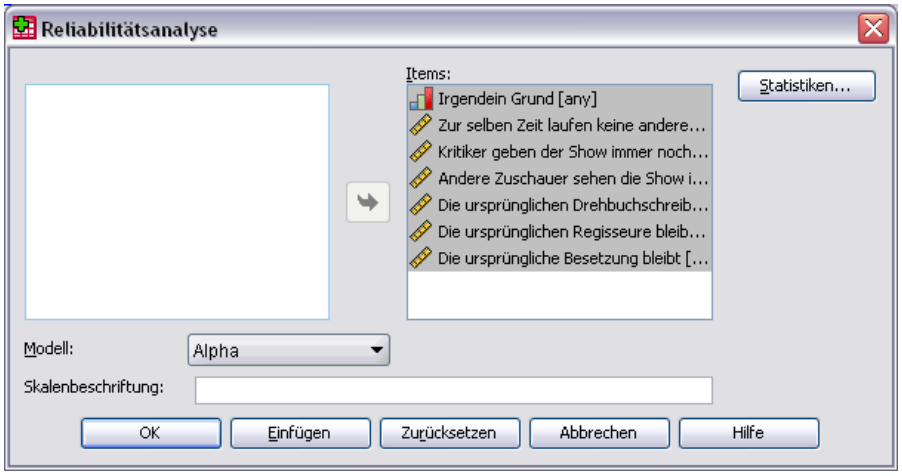
Wenn Sie homogene Variablengruppen identifizieren möchten, verwenden Sie die Prozedur “Hierarchische Clusteranalyse”, um Variablen zu clustern.

So lassen Sie eine Reliabilitätsanalyse berechnen:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Skalieren
  Reliabilitätsanalyse...

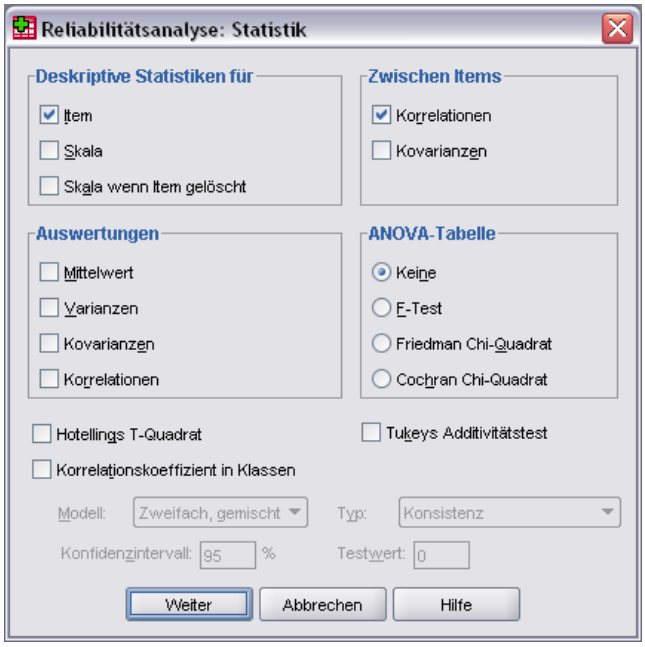
Abbildung 39-1
Dialogfeld “Reliabilitätsanalyse”



- Wählen Sie mindestens zwei Variablen als potentielle Komponenten einer additiven Skala aus.
- Wählen Sie aus dem Dropdown-Listenfeld “Modell” ein Modell aus.

Reliabilitätsanalyse: Statistik

Abbildung 39-2
Dialogfeld "Reliabilitätsanalyse: Statistik"



Sie können zahlreiche Statistiken auswählen, die sowohl die Skala als auch die Items beschreiben. Die Statistiken, die in der Standardeinstellung angezeigt werden, umfassen die Anzahl der Fälle, die Anzahl der Items und die folgenden Reliabilitätsschätzer:

- **Alpha-Modelle.** Bei dichotomen Daten entspricht dies dem Kuder-Richardson-20-(KR20-)Koeffizienten.
- **Split-Half-Modelle.** Korrelation zwischen den beiden Hälften, Split-Half-Reliabilität nach Guttman, Spearman-Brown-Reliabilität (gleiche und ungleiche Länge) und Alpha-Koeffizienten für jede Hälfte.
- **Guttman-Modelle.** Reliabilitätskoeffizienten Lambda 1 bis Lambda 6.
- **Parallele und streng parallele Modelle.** Anpassungstest für das Modell, Schätzer der Fehlervarianz, der Gesamtvarianz und der wahren Varianz, geschätzte gemeinsame Inter-Item-Korrelation, geschätzte Reliabilität und unverzerrter Schätzer der Reliabilität.

Deskriptive Statistiken für. Erzeugt deskriptive Statistiken für Skalen oder Items über Fälle.

- **Item.** Erzeugt deskriptive Statistiken für Items über Fälle.
- **Skala.** Erzeugt deskriptive Statistiken für Skalen.
- **Skala, wenn Item gelöscht.** Zeigt die Auswertungsstatistik an, bei der jedes Item mit der Skala verglichen wird, die aus den anderen Items aufgebaut wurde. Zu den statistischen Angaben gehören auch Mittelwert und Varianz der Skala, falls das Item aus der Skala gelöscht würde, die Korrelation zwischen dem Element und der Skala aus den anderen Items sowie Cronbachs Alpha, falls das Element aus der Skala gelöscht würde.

Auswertung. Hiermit werden deskriptive Statistiken der Item-Verteilungen für alle Items in der Skala berechnet.

- **Mittelwerte.** Auswertungsstatistik für die Mittelwerte der Items. Angezeigt werden der kleinste, größte und durchschnittliche Item-Mittelwert, der Bereich und die Varianz der Item-Mittelwerte sowie das Verhältnis zwischen dem größten und dem kleinsten Item-Mittelwert.
- **Varianzen.** Auswertungsstatistik für Varianzen der Items. Es werden die kleinsten, größten und mittleren Varianzen der Items, die Spannweite und die Varianz der Item-Varianzen sowie das Verhältnis zwischen den größten und den kleinsten Varianzen angezeigt.
- **Kovarianzen.** Statistik für die Kovarianzen zwischen den Items. Von den Kovarianzen zwischen den Items werden der kleinste und der größte Wert, der Mittelwert, die Spannweite und die Varianz sowie das Verhältnis vom größten zum kleinsten Wert angezeigt.
- **Korrelationen.** Statistik für die Korrelationen zwischen den Items. Von den Korrelationen zwischen den Items werden der kleinste und der größte Wert, der Mittelwert, die Spannweite und die Varianz sowie das Verhältnis vom größten zum kleinsten Wert angezeigt.

Inter-Item. Hiermit werden Matrizen der Korrelationen oder Kovarianzen zwischen den Items erstellt.

ANOVA-Tabelle. Hiermit werden Tests auf gleiche Mittelwerte berechnet.

- **F-Test.** Zeigt eine Tabelle zur Varianzanalyse mit Messwiederholungen an.
- **Friedman Chi-Quadrat.** Zeigt das Chi-Quadrat nach Friedman und den Konkordanz-Koeffizienten nach Kendall an. Diese Option ist für Daten geeignet, die in Form von Rängen vorliegen. Der Chi-Quadrat-Test ersetzt den üblichen F-Test in der ANOVA-Tabelle.
- **Cochran Chi-Quadrat.** Zeigt Cochrans Q-Test an. Diese Option ist für dichotome Daten geeignet. Die Q-Statistik ersetzt die übliche F-Statistik in der ANOVA-Tabelle.

Hotellings T-Quadrat. Erzeugt einen multivariaten Test der Nullhypothese, dass alle Items auf der Skala den gleichen Mittelwert besitzen.

Tukeys Additivitätstest. Erzeugt einen Test der Annahme, dass zwischen den Items keine multiplikative Wechselwirkung besteht.

Korrelationskoeffizienten in Klassen. Erzeugt ein Maß der Konsistenz oder Werteübereinstimmung innerhalb von Fällen.

- **Modell.** Wählen Sie das Modell für die Berechnung des Korrelationskoeffizienten in Klassen. Verfügbar sind die Modelle "Zwei-fach, gemischt", "Zwei-fach, zufällig" und "Ein-fach, zufällig". Wählen Sie Zwei-fach, gemischt aus, wenn die Personeneffekte zufällig und die Item-Effekte fest sind. Wählen Sie Zwei-fach, zufällig aus, wenn die Personeneffekte und die Item-Effekte zufällig sind. Wählen Sie Ein-fach, zufällig aus, wenn die Personeneffekte zufällig sind.
- **Typ.** Wählen Sie den Indextyp. "Konsistenz" und "Absolute Übereinstimmung" sind verfügbar.
- **Konfidenzintervall.** Legen Sie das Niveau des Konfidenzintervalls fest. Der Standardwert ist 95%.
- **Testwert.** Legen Sie den hypothetischen Wert des Koeffizienten für den Hypothesentest fest. Dies ist der Wert, mit dem der beobachtete Wert verglichen wird. Der Standardwert ist 0.

Zusätzliche Funktionen beim Befehl RELIABILITY

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Korrelationsmatrizen können gelesen und analysiert werden.
- Korrelationsmatrizen können für spätere Analysen gespeichert werden.
- Für die Split-Half-Methode können Aufteilungen festgelegt werden, die nicht genau Hälften entsprechen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
40

Multidimensionale Skalierung

Bei der multidimensionalen Skalierung wird versucht, die Struktur in einem Set von Distanzmaßen zwischen Objekten oder Fällen zu erkennen. Diese Aufgabe wird durch das Zuweisen von Beobachtungen zu bestimmten Positionen in einem konzeptuellen Raum (gewöhnlich zwei- oder dreidimensional) erzielt, und zwar so, dass die Distanzen zwischen den Punkten des Raums mit den gegebenen Unähnlichkeiten so gut wie möglich übereinstimmen. In vielen Fällen können die Dimensionen dieses konzeptuellen Raums interpretiert und für ein besseres Verständnis Ihrer Daten verwendet werden.

Wenn Sie über objektiv gemessene Variablen verfügen, können Sie die multidimensionale Skalierung als Technik zur Datenreduktion verwenden (erforderlichenfalls berechnet die Prozedur "Multidimensionale Skalierung" die Distanzen aus multivariaten Daten für Sie). Die multidimensionale Skalierung kann auch auf subjektive Einschätzungen von Unähnlichkeiten zwischen Objekten oder Konzepten angewendet werden. Außerdem kann die Prozedur "Multidimensionale Skalierung" Unähnlichkeitsdaten aus mehreren Quellen verarbeiten, beispielsweise von mehreren Befragern oder Befragten einer Umfrage.

Beispiel. Wie nehmen Personen Ähnlichkeiten zwischen unterschiedlichen Autos wahr? Wenn Sie über Daten verfügen, in denen Befragte ihre Einschätzungen der Ähnlichkeiten von verschiedenen Automarken und -modellen abgegeben haben, kann die multidimensionale Skalierung zur Identifizierung der Dimensionen verwendet werden, welche die Wahrnehmungen von Käufern beschreibt. Sie könnten zum Beispiel feststellen, dass Preis und Größe eines Fahrzeuges einen zweidimensionalen Raum definieren, welcher die von den Befragten geäußerten Ähnlichkeiten erklärt.

Statistiken. Für jedes Modell: Datenmatrix, optimal skalierte Datenmatrix, S-Stress (Young), Stress (Kruskal), RSQ, Stimulus-Koordinaten, durchschnittlicher Stress und RSQ für jeden Stimulus (RMDS-Modelle). Für Modelle der individuellen Differenzen (INDSCAL): Subjektgewichtungen und Seltsamkeits-Index ("weirdness index") für jedes Subjekt. Für jede Matrix in replizierten Modellen für die multidimensionale Skalierung: Stress und RSQ für jeden Stimulus. Diagramme: Stimulus-Koordinaten (zwei- oder dreidimensional), Streudiagramm von Unähnlichkeiten über Distanzen.

Daten. Wenn Sie über Unähnlichkeitsdaten verfügen, sollten alle Unähnlichkeiten quantitativ und mit derselben Maßeinheit gemessen sein. Wenn Sie über multivariate Daten verfügen, können die Variablen quantitativ, binär oder Häufigkeitsdaten sein. Die Skalierung der Variablen ist ein wichtiger Punkt. Unterschiede in der Skalierung können Ihre Lösung beeinflussen. Wenn Ihre Variablen große Differenzen in der Skalierung aufweisen (wenn zum Beispiel eine Variable in Dollar und die andere Variable in Jahren gemessen wird), sollten Sie deren Standardisierung in Betracht ziehen (dies kann mit der Prozedur "Multidimensionale Skalierung" automatisch durchgeführt werden).

Annahmen. Die Prozedur “Multidimensionale Skalierung” ist relativ frei von Annahmen zur Verteilung. Stellen Sie sicher, dass Sie im Dialogfeld “Multidimensionale Skalierung: Optionen” ein geeignetes Messniveau auswählen (Ordinal-, Intervall- oder Verhältnisdaten), sodass Ihre Ergebnisse richtig berechnet werden können.

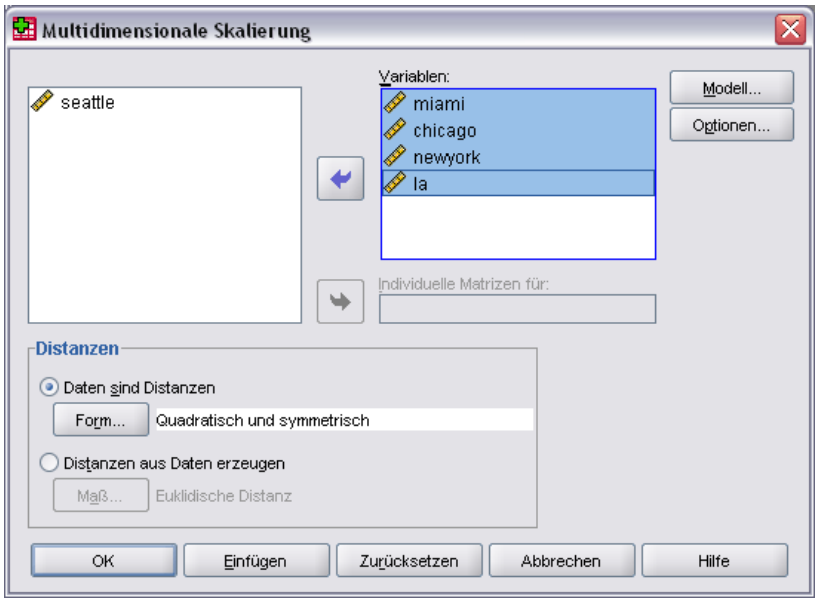
Verwandte Prozeduren. Wenn Sie eine Datenreduktion durchführen möchten, können Sie auch eine Faktoranalyse durchführen, insbesondere bei quantitativen Variablen. Wenn Sie Gruppen von ähnlichen Fällen identifizieren möchten, können Sie die multidimensionale Skalierung durch eine hierarchische Clusteranalyse oder eine Clusterzentrenanalyse ergänzen.

So berechnen Sie eine multidimensionale Skalierung:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Skalieren
    Multidimensionale Skalierung...

Abbildung 40-1
Dialogfeld “Multidimensionale Skalierung”



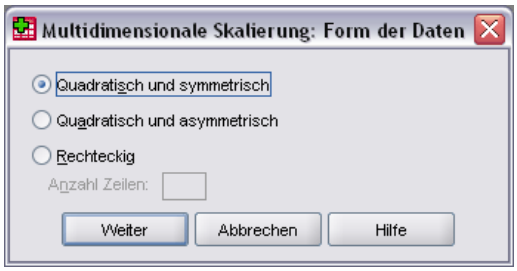
- Wählen Sie unter “Distanzen” entweder Daten sind Distanzen oder Distanzen aus Daten erzeugen aus.

- Wenn Ihre Daten Distanzen darstellen, wählen Sie mindestens vier numerische Variablen für die Analyse aus. (Sie können auch auf Form klicken, um die Form der Distanzmatrix anzugeben.)

- Wenn Sie die Distanzen erstellen möchten, bevor diese analysiert werden, müssen Sie mindestens eine numerische Variable auswählen. (Sie können auch auf Maß klicken, um den Typ des gewünschten Distanzmaßes anzugeben.) Jede Gruppenvariable kann entweder numerisch oder ein String sein und Sie können getrennte Matrizen für jede Kategorie einer Gruppenvariablen erstellen, indem Sie diese Variable in die Liste “Individuelle Matrizen für” verschieben.

Multidimensionale Skalierung: Form der Daten

Abbildung 40-2
Dialogfeld "Multidimensionale Skalierung: Form der Daten"

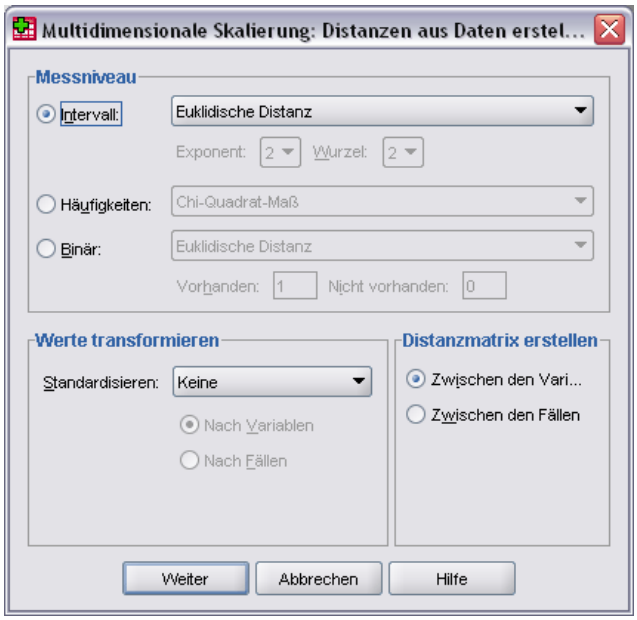


Wenn die Arbeitsdatei Distanzen innerhalb einer Gruppe von Objekten oder zwischen zwei Gruppen von Objekten darstellt, müssen Sie die Form der Datenmatrix angeben, um die richtigen Ergebnisse zu erhalten.

Anmerkung: Sie können Quadratisch und symmetrisch nicht auswählen, wenn im Dialogfeld "Modell" eine Konditionalität der Zeilen festgelegt ist.

Multidimensionale Skalierung: Distanzen aus Daten erstellen

Abbildung 40-3
Dialogfeld "Multidimensionale Skalierung: Distanzen aus Daten erstellen"



Die multidimensionale Skalierung verwendet Unähnlichkeitsdaten, um eine Skalierungslösung zu erstellen. Wenn Ihre Daten multivariate Daten darstellen (Werte gemessener Variablen), müssen Sie Unähnlichkeitsdaten erstellen, um eine multidimensionale Skalierungslösung berechnen zu können. Sie können Optionen für das Erstellen von Unähnlichkeitsmaßen aus Ihren Daten festlegen.

Messniveau. Hier können Sie das Unähnlichkeitsmaß für Ihre Analyse festlegen. Wählen Sie im Gruppenfeld "Maß" die Option aus, die Ihrem Datentyp entspricht. Wählen Sie dann aus dem Dropdown-Listenfeld ein Maß aus, das diesem Messwerttyp entspricht. Die folgenden Optionen sind verfügbar:

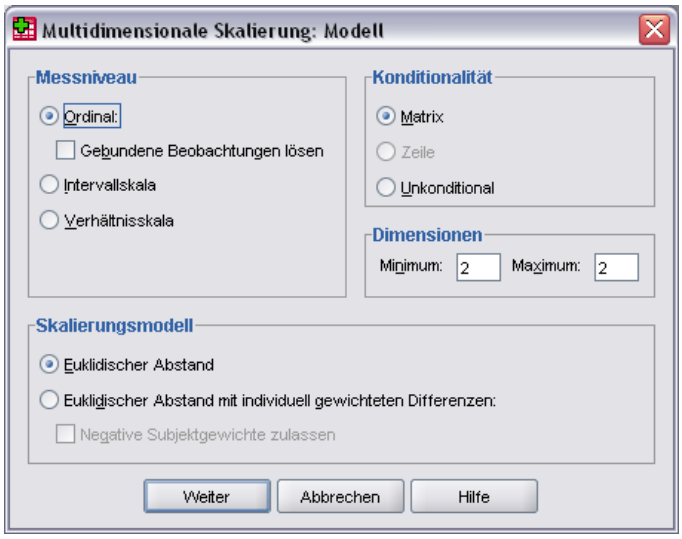
- **Intervall**. Euklidischer Abstand, quadrierter Euklidischer Abstand, Tschebyscheff, Block, Minkowski oder ein benutzerdefiniertes Maß.
- **Häufigkeiten**. Chi-Quadrat-Maß oder Phi-Quadrat-Maß.
- **Binär**. Euklidischer Abstand, quadrierter Euklidischer Abstand, Größendifferenz, Musterdifferenz, Varianz und Distanzmaß nach Lance und Williams.

Distanzmatrix erstellen. Mit dieser Funktion können Sie die Einheit der Analyse wählen. Zur Auswahl stehen "Zwischen den Variablen" oder "Zwischen den Fällen".

Werte transformieren. In bestimmten Fällen, zum Beispiel wenn die Variablen mit sehr unterschiedlichen Skalen gemessen werden, empfiehlt sich das Standardisieren der Werte vor dem Berechnen der Ähnlichkeiten (nicht auf binäre Daten anwendbar). Wählen Sie in der Dropdown-Liste "Standardisieren" eine Standardisierungsmethode aus. Wenn keine Standardisierung erforderlich ist, wählen Sie Keine aus.

Multidimensionale Skalierung: Modell

Abbildung 40-4
Dialogfeld "Multidimensionale Skalierung: Modell"



Die richtige Schätzung eines Modells für die multidimensionale Skalierung hängt von Aspekten der Daten und dem Modell selbst ab.

Messniveau. Mit dieser Funktion können Sie das Niveau Ihrer Daten festlegen. Die Optionen "Ordinalskala", "Intervallskala" und "Verhältnisskala" sind verfügbar. Wenn die Variablen ordinal sind, können Sie Gebundene Beobachtungen lösen auswählen. Die Variablen werden dann wie stetige Variablen behandelt, sodass die Bindungen (gleiche Werte für unterschiedliche Fälle) optimal gelöst werden können.

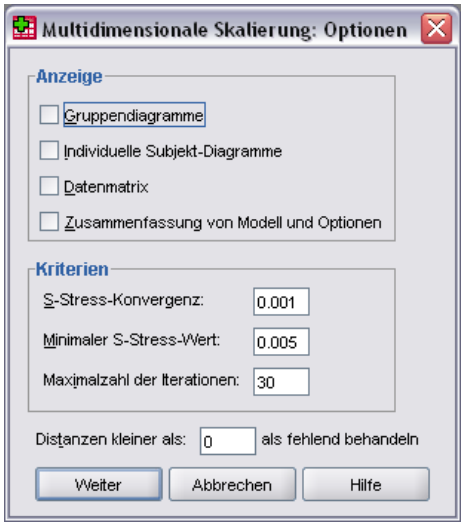
Konditionalität. Hiermit können sie festlegen, welche Vergleiche sinnvoll sind. Als Optionen sind "Matrix", "Zeile" und "Unkonditional" verfügbar.

Dimensionen. Mit dieser Funktion können Sie die Anzahl der Dimensionen für die Skalierungslösung(en) festlegen. Für jede Zahl im Bereich wird eine Lösung berechnet. Legen Sie ganze Zahlen zwischen 1 und 6 fest. Ein Minimum von 1 ist nur möglich, wenn Sie als Skalierungsmodell Euklidischer Abstand auswählen. Legen Sie die gleiche Zahl für das Minimum und das Maximum fest, wenn Sie nur eine Lösung wünschen.

Skalierungsmodell. Hiermit können Sie die Annahmen festlegen, nach denen die Skalierung durchgeführt wird. Als Optionen sind "Euklidischer Abstand" oder "Euklidischer Abstand mit individuell gewichteten Differenzen" (auch als INDSCAL bekannt) verfügbar. Beim Modell "Euklidischer Abstand mit individuell gewichteten Differenzen" können Sie Negative Subjektgewichte zulassen auswählen, wenn dies für Ihre Daten geeignet ist.

Multidimensionale Skalierung: Optionen

Abbildung 40-5
Dialogfeld "Multidimensionale Skalierung: Optionen"



Sie können Optionen für die Analyse der multidimensionalen Skalierung festlegen.

Anzeigen. Mit dieser Funktion können Sie verschiedene Ausgabetypen auswählen. Die Optionen "Gruppendiagramme", "Individuelle Subjekt-Diagramme", "Datenmatrix" und "Zusammenfassung von Modell und Optionen" sind verfügbar.

Kriterien. Hiermit können Sie bestimmen, wann die Iterationen beendet werden sollen. Um die Standardeinstellungen zu ändern, geben Sie Werte für S-Stress-Konvergenz, Minimaler S-Stress-Wert und Iterationen, max. ein.

Distanzen kleiner n als fehlend behandeln. Distanzen, die einen geringeren Wert als diesen Wert aufweisen, werden aus der Analyse ausgeschlossen.

Zusätzliche Funktionen beim Befehl ALSCAL

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Es können drei weitere Modelltypen verwendet werden. Diese sind in der Literatur über die multidimensionale Skalierung als ASCAL, AINDS und GEMSCAL bekannt.
- Es können polynomiale Transformationen von Intervall- und Verhältnisdaten ausgeführt werden.
- Bei ordinalen Daten können statt Distanzen Ähnlichkeiten analysiert werden.
- Es können nominale Daten analysiert werden.
- Verschiedene Koordinatenmatrizen und Gewichtungsmatrizen können in Dateien gespeichert und für eine Analyse erneut eingelesen werden.
- Die multidimensionale Entfaltung kann eingeschränkt werden.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
41

Verhältnisstatistik

Die Prozedur “Verhältnisstatistik” bietet eine umfassende Liste mit Auswertungsstatistiken zur Beschreibung des Verhältnisses zwischen zwei metrischen Variablen.

Sie können die Ausgabe nach Werten einer Gruppenvariablen in auf- oder absteigender Reihenfolge sortieren. Der Bericht für die Verhältnisstatistik kann in der Ausgabe unterdrückt werden, und die Ergebnisse können in einer externen Datei gespeichert werden.

Beispiel. Ist das Verhältnis zwischen dem Schätzwert und dem Verkaufspreis von Häusern in fünf Verwaltungsbezirken in etwa gleich? Im Ergebnis der Analyse könnte sich herausstellen, dass die Verteilung der Verhältnisse je nach Bezirk erheblich variiert.

Statistiken. Median, Mittel, gewichtetes Mittel, Konfidenzintervalle, Streuungskoeffizient (COD), medianzentrierter Variationskoeffizient, mittelzentrierter Variationskoeffizient, preisbezogenes Differential (PRD), Standardabweichung, durchschnittliche absolute Abweichung (AAD), Bereich, Mindest- und Höchstwerte sowie der Konzentrationsindex, der für einen benutzerdefinierten Bereich oder Prozentsatz innerhalb des Medianverhältnisses berechnet wird.

Daten. Verwenden Sie zum Kodieren von Gruppenvariablen (nominales oder ordinales Messniveau) numerische Codes oder Strings

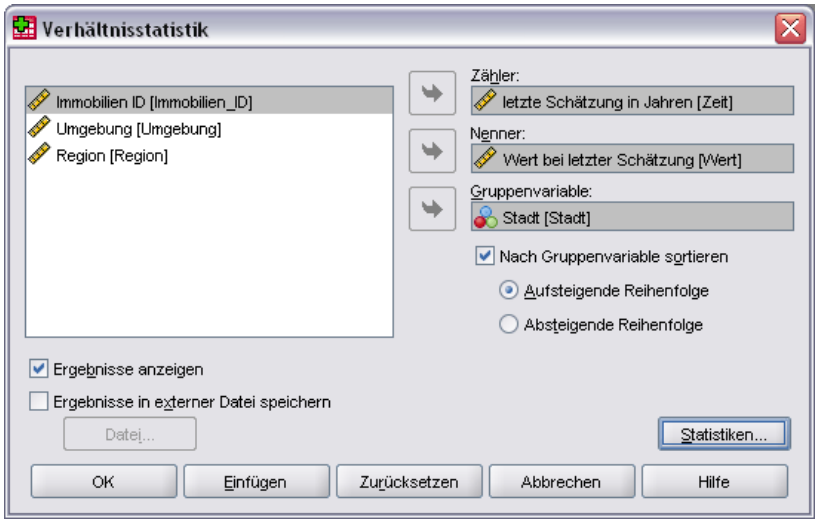
Annahmen. Die Variablen, durch die Zähler und Nenner des Verhältnisses definiert werden, müssen metrische Variablen sein, die positive Werte akzeptieren.

So lassen Sie Verhältnisstatistiken berechnen:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Deskriptive Statistiken
    Verhältnis...

Abbildung 41-1
Dialogfeld "Verhältnisstatistik: Statistik"



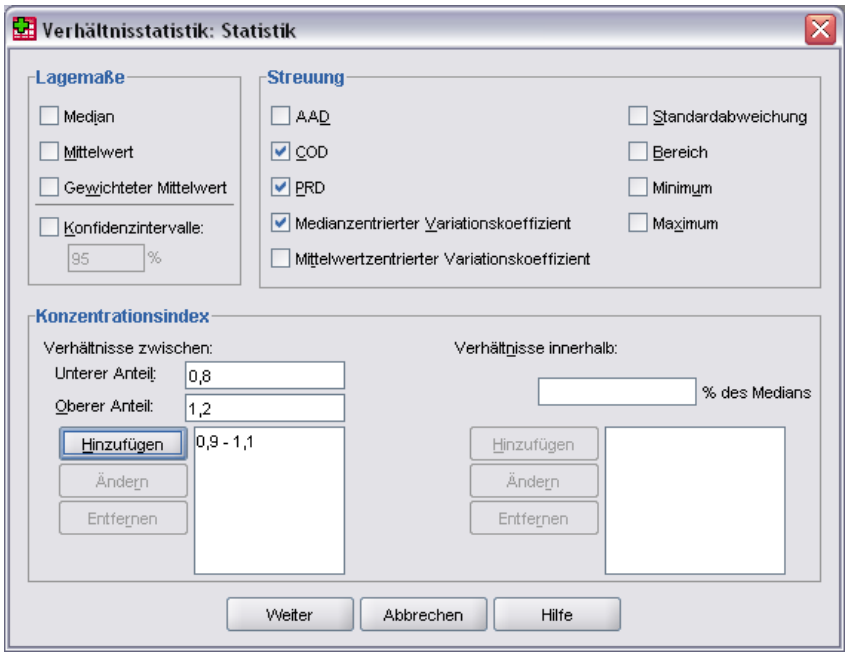
- ▶ Wählen Sie eine Zählervariable.
- ▶ Wählen Sie eine Nennervariable.

Die folgenden Optionen sind verfügbar:

- Wählen Sie eine Gruppenvariable, und legen Sie die Reihenfolge der Gruppen in den Ergebnissen fest.
- Wählen Sie aus, ob die Ergebnisse im Viewer angezeigt werden sollen.
- Legen Sie fest, ob die Ergebnisse zur späteren Verwendung in einer externen Datei gespeichert werden sollen, und geben Sie einen Namen für diese Datei an.

Verhältnisstatistik

Abbildung 41-2
Dialogfeld "Verhältnisstatistik: Statistik"



Lagemaße. Lagemaße sind Statistiken, mit denen die Verteilung von Verhältnissen beschrieben wird.

- **Median.** Der Wert, der sich ergibt, wenn die Anzahl der Verhältnisse unterhalb dieses Werts gleich der Anzahl der Verhältnisse oberhalb dieses Werts ist.
- **Mittelwert.** Das Ergebnis aus der Summierung aller Verhältnisse und der anschließenden Division des Ergebnisses durch die Gesamtanzahl der Verhältnisse.
- **Gewichteter Mittelwert.** Das Ergebnis aus der Division des Mittelwerts für den Zähler durch den Mittelwert für den Nenner. Der gewichtete Mittelwert ist außerdem der Mittelwert der durch den Nenner gewichteten Verhältnisse.
- **Konfidenzintervalle.** Mit dieser Option werden Konfidenzintervalle für den Mittelwert, den Median und den gewichteten Mittelwert (falls gewünscht) angezeigt. Geben Sie für das Konfidenzniveau einen Wert größer oder gleich 0 und kleiner als 100 ein.

Streuung. Statistiken, mit denen die Variation oder Streubreite in den beobachteten Werten gemessen wird.

- **AAD.** Die durchschnittliche absolute Abweichung ist die Summe aus den absoluten Abweichungen der Verhältnisse des Medians und der Division des Ergebnisses durch die Gesamtanzahl der Verhältnisse.
- **COD.** Der Streuungskoeffizient entspricht der durchschnittlichen absoluten Abweichung in Prozent des Medians.
- **PRD.** Das preisbezogene Differential, auch Index der Regressivität genannt, ist das Ergebnis der Division des Mittelwerts durch den gewichteten Mittelwert.

- **Medianzentrierter Variationskoeffizient.** Der medianzentrierte Variationskoeffizient entspricht der Wurzel der mittleren quadratischen Abweichung vom Median in Prozent des Medians.
- **Mittelwertzentrierter Variationskoeffizient.** Der mittelwertzentrierte Variationskoeffizient entspricht der Standardabweichung in Prozent des Mittelwerts.
- **Standardabweichung.** Die Standardabweichung ist das Ergebnis der Summierung der quadratischen Abweichungen der Verhältnisse zum Mittelwert, der Division des Ergebnisses durch die Gesamtanzahl der Verhältnisse minus eins und der Berechnung der positiven Quadratwurzel.
- **Spannweite.** Die Spannweite ist das Ergebnis der Subtraktion des minimalen Verhältnisses vom maximalen Verhältnis.
- **Minimum.** Das Minimum ist das kleinste Verhältnis.
- **Maximum.** Das Maximum ist das größte Verhältnis.

Konzentrationsindex. Der Konzentrationskoeffizient mißt den prozentualen Anteil der Verhältnisse, die in einem bestimmten Intervall liegen. Dieser Koeffizient kann auf zwei verschiedene Arten berechnet werden:

- **Verhältnisse zwischen.** Bei dieser Option wird das Intervall explizit durch Angabe der unteren und oberen Intervallwerte definiert. Geben Sie Werte für den unteren Anteil und den oberen Anteil ein und klicken Sie auf Hinzufügen, um ein Intervall auszugeben.
- **Verhältnisse innerhalb.** Bei dieser Option wird das Intervall implizit durch Angabe des prozentualen Medians definiert. Geben Sie einen Wert zwischen 0 und 100 ein und klicken Sie auf Hinzufügen. Die untere Grenze des Intervalls ist gleich (1 – 0,01 × Wert) × Median. Die obere Grenze ist gleich (1 + 0,01 × Wert) × Median.

Kapitel
42

ROC-Kurven

Diese Prozedur stellt einen sinnvollen Weg zur Beurteilung von Klassifikationsschemata dar, bei denen eine Variable mit zwei Kategorien verwendet wird, um Subjekte zu klassifizieren.

Beispiel. Es liegt im Interesse von Banken, Kunden ordnungsgemäß danach zu klassifizieren, ob diese Kunden mit ihren Darlehen in Verzug geraten werden oder nicht. Daher werden spezielle Verfahren für diese Entscheidungen entwickelt. Mithilfe von ROC-Kurven kann beurteilt werden, wie gut diese Verfahren funktionieren.

Statistiken. Fläche unter der ROC-Kurve mit Konfidenzintervall und Koordinaten-Punkten der ROC-Kurve. Diagramme: ROC-Kurve.

Methoden. Die Schätzung der Fläche unter der ROC-Kurve kann parameterunabhängig oder parameterabhängig unter Verwendung eines binegativ exponentiellen Modells erfolgen.

Daten. Die Testvariablen sind quantitativ. Die Testvariablen setzen sich oft aus Wahrscheinlichkeiten aus der Diskriminanzanalyse bzw. logistischen Regression zusammen oder sie werden aus Werten auf einer willkürlichen Skala zusammengesetzt, die anzeigen, wie sehr ein Beurteiler davon "überzeugt" ist, dass ein Subjekt in die eine oder die andere Kategorie fällt. Der Typ der Zustandsvariablen ist nicht vorgegeben. Diese Variable zeigt die tatsächliche Kategorie an, zu der ein Subjekt gehört. Der Wert der Zustandsvariablen zeigt an, welche Kategorie als *positiv* zu betrachten ist.

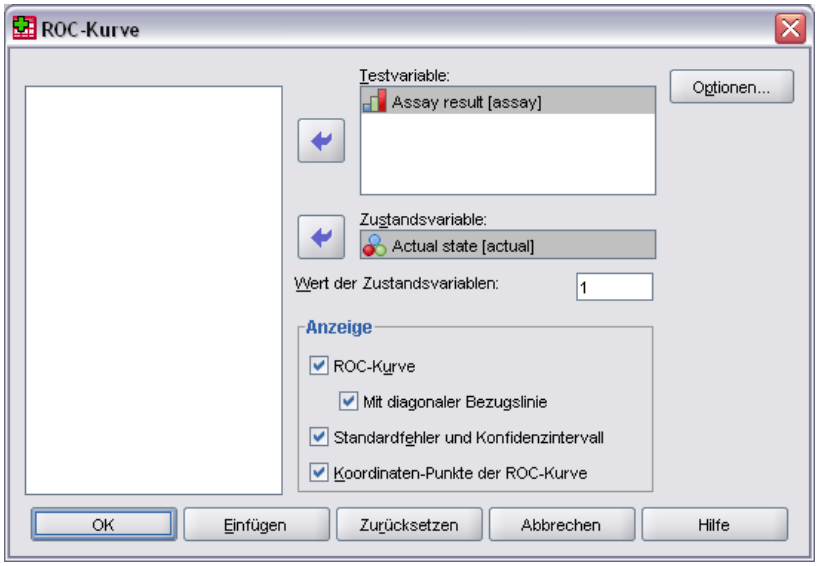
Annahmen. Es wird angenommen, dass ansteigende Werte auf der Skala des Beurteilers ein Ansteigen der Überzeugung darstellen, dass das Subjekt in die eine Kategorie fällt. Abfallende Werte auf der Skala stellen hingegen eine ansteigende Überzeugung dar, dass das Subjekts der anderen Kategorie angehört. Der Anwender wählt aus, welche Richtung als *positiv* anzusehen ist. Es wird außerdem angenommen, dass die *tatsächliche* Kategorie bekannt ist, zu der jedes Subjekt gehört.

So Erstellen Sie eine ROC-Kurve:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  ROC-Kurve...

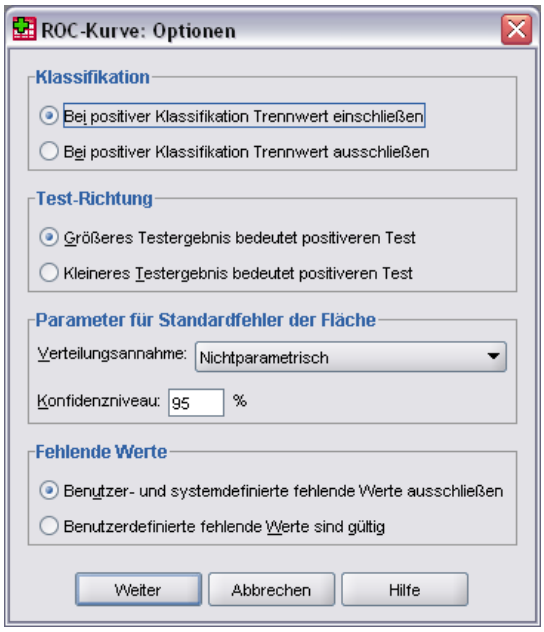
Abbildung 42-1
Dialogfeld "ROC-Kurve"



- Wählen Sie mindestens eine Wahrscheinlichkeitsvariable für den Test aus.
- Wählen Sie eine Zustandsvariable aus.
- Legen Sie den *positiven* Wert für die Zustandsvariable fest.

ROC-Kurve: Optionen

Abbildung 42-2
Dialogfeld "ROC-Kurve: Optionen"



Sie können eine der folgenden Optionen für die ROC-Analyse auswählen:

Klassifikation. Hiermit können Sie festlegen, ob der Trennwert bei einer *positiven* Klassifikation einbezogen oder ausgeschlossen werden soll. Diese Einstellung hat gegenwärtig keine Auswirkungen auf die Ausgabe.

Test-Richtung. Hiermit geben Sie die Richtung der Skala bezogen auf die *positive* Kategorie an.

Parameter für Standardfehler der Fläche. Hiermit geben Sie die Methode an, mit welcher der Standardfehler der Fläche unter der Kurve geschätzt wird. Es stehen eine nichtparametrische und eine binegative exponentielle Methode zur Verfügung. Sie können hier außerdem das Niveau des Konfidenzintervalls festlegen. Es sind Werte zwischen 50,1% und 99,9% möglich.

Fehlende Werte. Hier können Sie festlegen, wie fehlende Werte behandelt werden.

Kapitel
43

Übersicht über die Diagrammfunktion

Diagramme mit hoher Auflösung können mit den Verfahren im Menü “Grafiken” und mit etlichen der Verfahren im Menü “Analysieren” erstellt werden. In diesem Kapitel finden Sie eine Übersicht über die Diagrammfunktion.

Erstellen und Ändern von Diagrammen

Bevor Sie ein Diagramm erstellen können, müssen Sie über Daten in Ihrem Daten-Editor verfügen. Sie können die Daten direkt in den Daten-Editor eingeben, eine bereits gespeicherte Datendatei öffnen oder ein Arbeitsblatt, eine durch Tabulatoren getrennte Datendatei oder eine Datenbankdatei einlesen. In der Menüauswahl “Lernprogramm” im Menü “Hilfe” finden Sie Online-Beispiele zum Erstellen und Ändern von Diagrammen, und das Online-Hilfesystem bietet Informationen darüber, wie alle Diagrammtypen erstellt und geändert werden können.

Erstellen von Diagrammen

In der Diagrammerstellung können Sie aus vordefinierten Galeriediagrammen oder aus einzelnen Bestandteilen (z. B. Achsen und Balken) Diagramme erstellen. Sie erstellen ein Diagramm, indem Sie die Galeriediagramme bzw. die Grundelemente auf die **Zeichenfläche** ziehen. Dies ist die große Fläche rechts neben der Liste “Variablen” im Dialogfeld “Diagrammerstellung”.

Während Sie das Diagramm erstellen, wird auf der Zeichenfläche eine Vorschau des Diagramms angezeigt. Obwohl die Vorschau vordefinierte Variablenlabels und Messniveaus verwendet, werden nicht die tatsächlichen Daten angezeigt. Stattdessen werden Zufallsdaten verwendet, um einen Rohentwurf des Diagramms anzufertigen.

Für neue Benutzer bietet sich die Verwendung der Galerie an. Informationen zur Verwendung der Galerie finden Sie unter Erstellen eines Diagramms aus der Galerie auf S. 475.

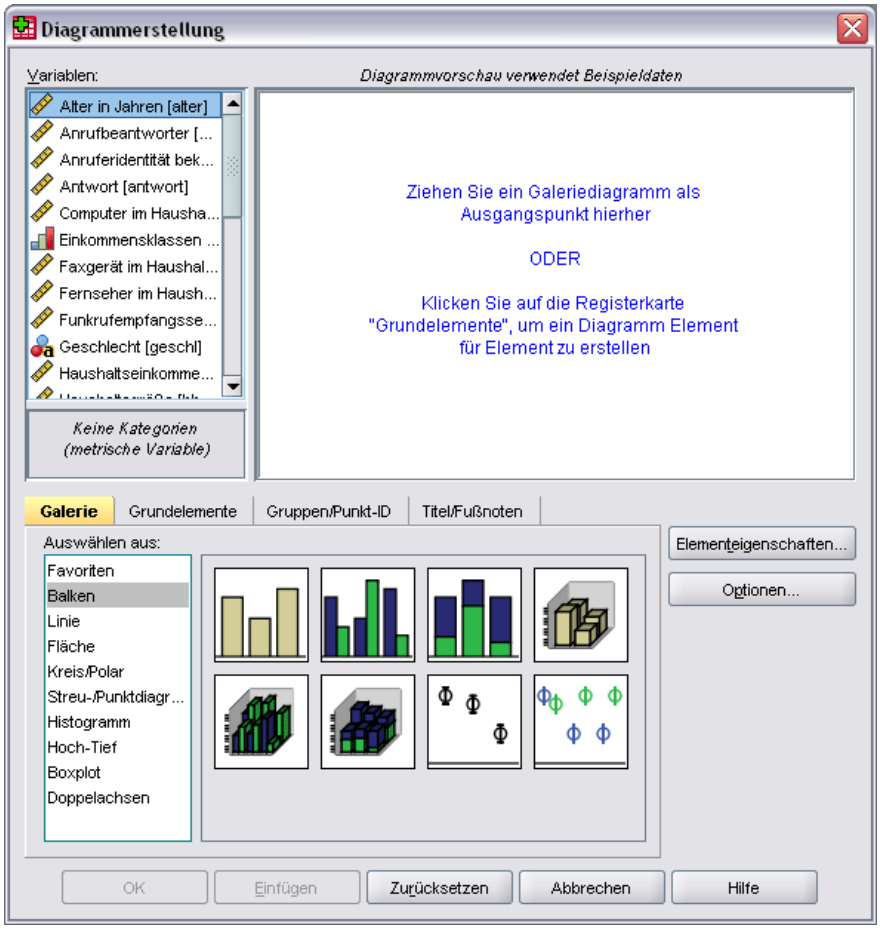
So starten Sie die Diagrammerstellung:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Grafiken
  Diagrammerstellung

Damit wird das Dialogfeld “Diagrammerstellung” geöffnet.

Abbildung 43-1
Dialogfeld "Diagrammerstellung"



Erstellen eines Diagramms aus der Galerie

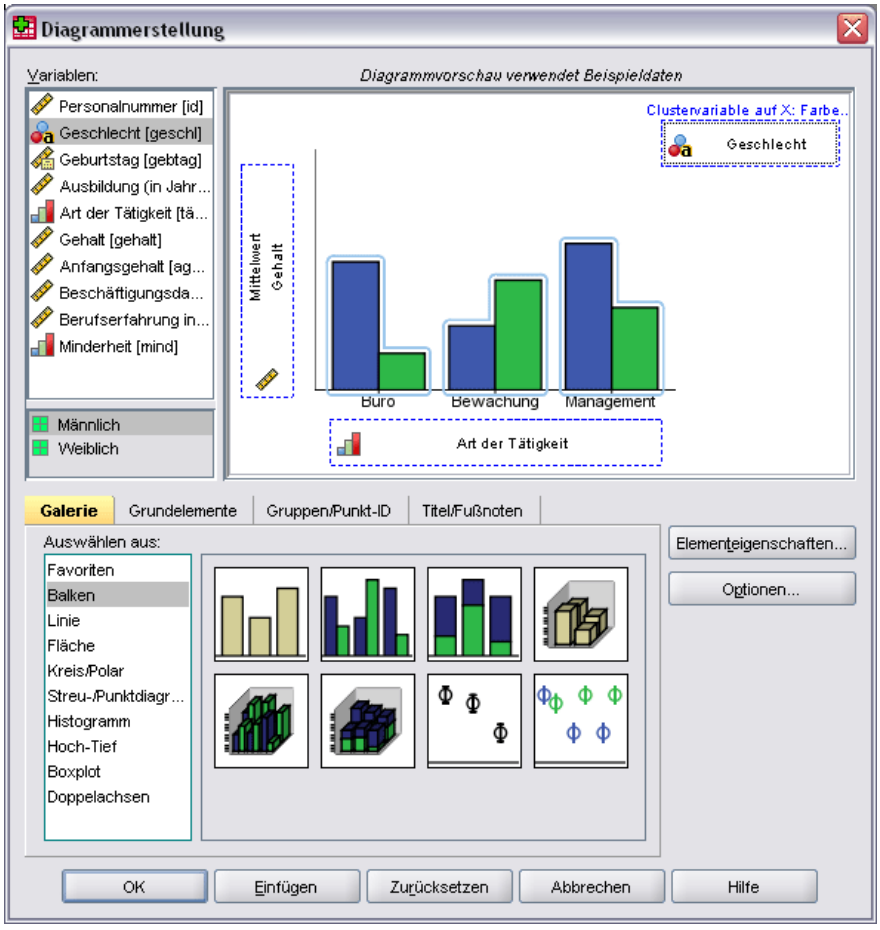
Die einfachste Methode, ein Diagramm zu erstellen, ist die Verwendung der Galerie. Im Folgenden finden Sie allgemeine Schritte zum Erstellen eines Diagramms aus der Galerie.

- Klicken Sie auf die Registerkarte Galerie, falls diese nicht bereits angezeigt wird.
- Wählen Sie in der Liste "Auswählen aus" eine Diagrammkategorie aus. Jede Kategorie enthält mehrere Typen.
- Ziehen Sie das Bild des gewünschten Diagramms auf die Zeichenfläche. Sie können auch auf das Bild doppelklicken. Wenn in der Zeichenfläche bereits ein Diagramm angezeigt wird, ersetzt das Galeriediagramm das Achsenset und die Grafikelemente im Diagramm.
- Sie können Variablen aus der Liste "Variablen" in die Achsenablagezonen und, sofern verfügbar, in die Gruppierungsablagezone ziehen. Wenn in einer Achsenablagezone bereits eine Statistik angezeigt wird und Sie die Statistik verwenden möchten, müssen Sie keine Variable in die

Ablagezone ziehen. Sie müssen der Zone nur dann eine Variable hinzufügen, wenn der Text in der Zone blau ist. Wenn der Text schwarz ist, enthält die Zone bereits eine Variable oder eine Statistik.

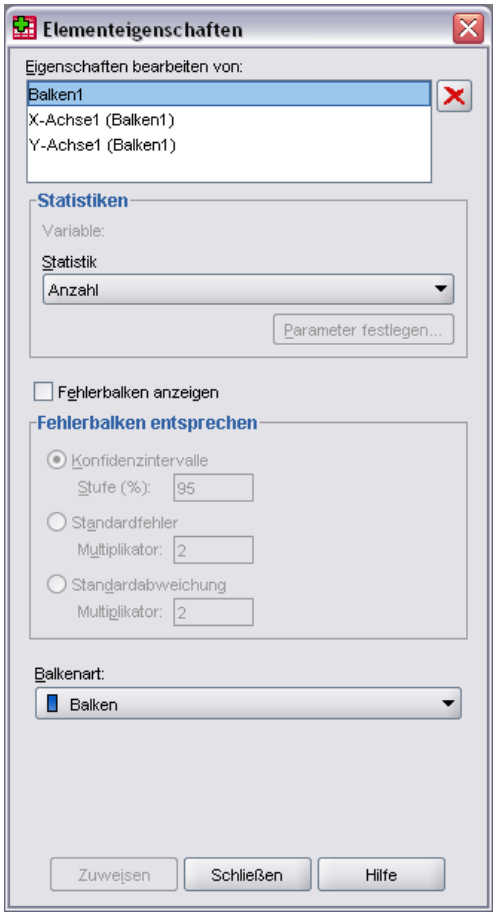
Anmerkung: Das Messniveau der Variablen ist wichtig. Während Sie das Diagramm erstellen, werden in der Diagrammerstellung Standardeinstellungen auf der Grundlage des Messniveaus festgelegt. Darüber hinaus kann das resultierende Diagramm für verschiedene Messniveaus durchaus unterschiedlich aussehen. Sie können das Messniveau einer Variablen vorübergehend ändern, indem Sie mit der rechten Maustaste auf die Variable klicken und eine Option auswählen.

Abbildung 43-2
Dialogfeld "Diagrammerstellung" mit belegten Ablagezonen



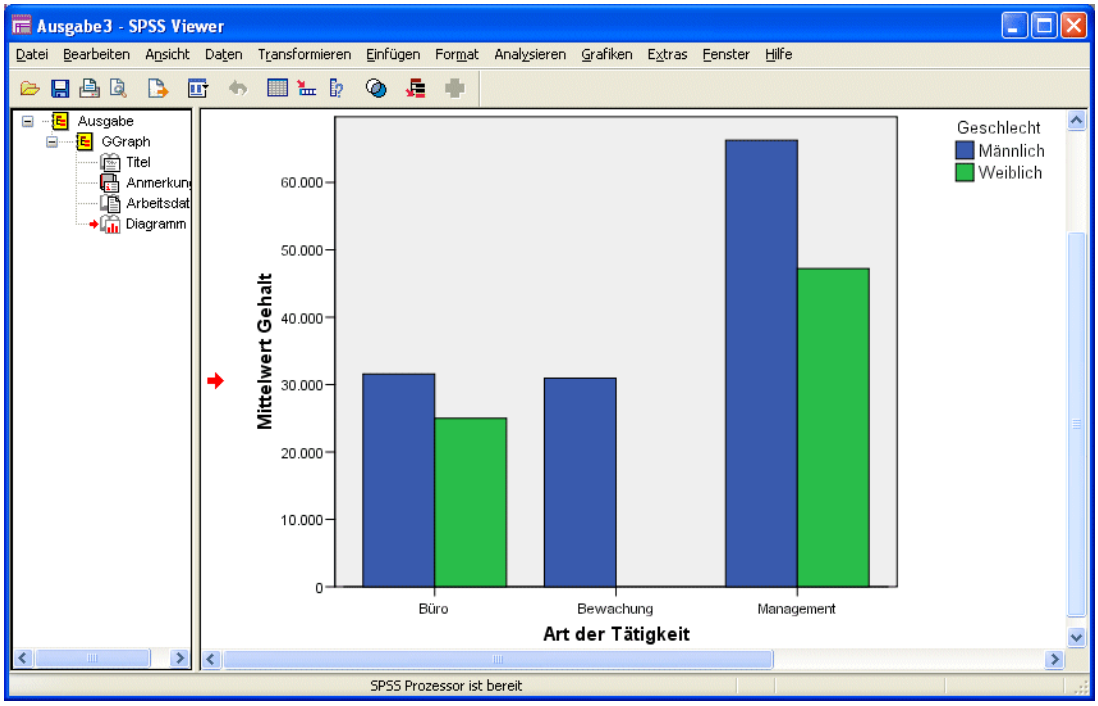
- Wenn Sie die Statistiken oder die Attribute für die Achsen oder Legenden ändern möchten (z. B. den Skalenbereich), klicken Sie auf Elementeigenschaften.

Abbildung 43-3
Dialogfeld "Elementeigenschaften"



- Wählen Sie in der Liste "Eigenschaften bearbeiten von" das Element aus, das Sie bearbeiten möchten. (Wenn Sie Informationen zu den einzelnen Eigenschaften benötigen, klicken Sie auf Hilfe.)
- Klicken Sie nach dem Vornehmen von Änderungen auf Zuweisen.
- Wenn Sie mehrere Variablen zu dem Diagramm hinzufügen müssen (beispielsweise für Clustering oder Feldeinteilung), klicken Sie auf die Registerkarte Gruppen/Punkt-ID im Dialogfeld "Diagrammerstellung" und wählen Sie eine oder mehrere Optionen aus. Ziehen Sie anschließend kategoriale Variablen in die neuen Ablagezonen, die in der Zeichenfläche angezeigt werden.
- Wenn Sie das Diagramm transponieren möchten (um beispielsweise die Balken horizontal auszurichten), klicken Sie auf die Registerkarte Grundelemente und anschließend auf Transponieren.
- Klicken Sie auf OK, um das Diagramm zu erstellen. Das Diagramm wird im Viewer angezeigt.

Abbildung 43-4
Balkendiagramm im Viewer-Fenster



Bearbeiten von Diagrammen

Der Diagramm-Editor bietet eine leistungsstarke, benutzerfreundliche Umgebung, in der Sie Ihre Diagramme anpassen und eine explorative Datenanalyse an Ihren Daten vornehmen können. Der Diagramm-Editor bietet folgende Funktionen:

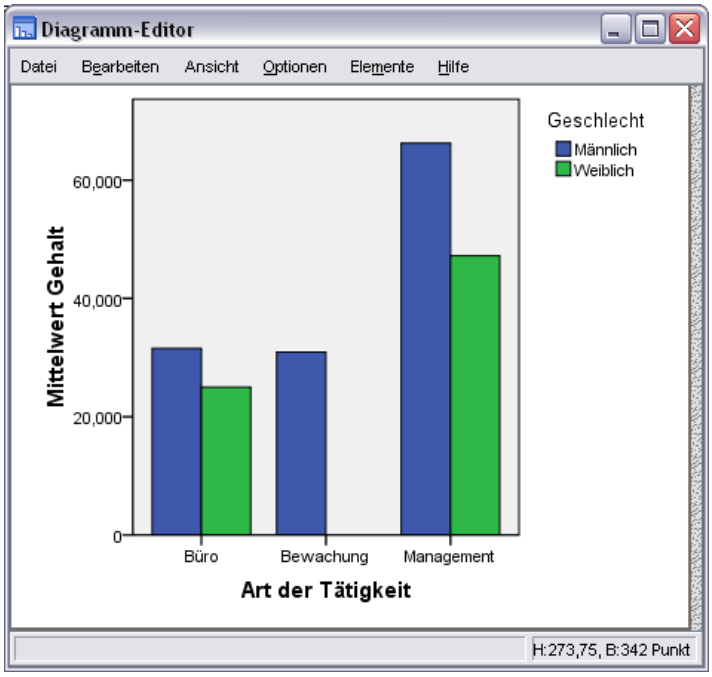
- **Einfache und intuitive Benutzeroberfläche.** Mithilfe von Menüs, Kontextmenüs und Symbolleisten können Sie schnell und einfach Teile des Diagramms auswählen und bearbeiten. Sie können Texte auch direkt in Diagramme eingeben.
- **Große Bandbreite an Formatierungs- und Statistikoptionen.** Sie können aus einer Vielzahl von Stilen und Statistikoptionen auswählen.
- **Leistungsstarke Untersuchungswerkzeuge.** Sie können an Ihren Daten auf verschiedene Weise eine explorative Datenanalyse durchführen, beispielsweise durch Beschriften, Ändern der Reihenfolge und durch Drehen. Sie können den Diagrammtyp und die Rollen der Variablen im Diagramm ändern. Außerdem können Sie Verteilungskurven sowie Anpassungs-, Interpolations- und Bezugslinien einfügen.
- **Flexible Vorlagen für konsistentes Erscheinungsbild und Verhalten.** Sie können benutzerdefinierte Vorlagen erstellen und damit spielend Diagramme anfertigen, die das von Ihnen gewünschte Erscheinungsbild und die gewünschten Optionen enthalten. Beispiel: Wenn die Achsenbeschriftungen immer eine bestimmte Ausrichtung haben sollen, können Sie die Ausrichtung in einer Vorlage festlegen und die Vorlage auf andere Diagramme anwenden.

So lassen Sie den Diagramm-Editor anzeigen:

- Erstellen Sie ein Diagramm in SPSS oder öffnen Sie eine Viewer-Datei mit Diagrammen.
- Doppelklicken Sie im Viewer auf ein Diagramm.

  Dadurch wird das Diagramm im Diagramm-Editor angezeigt.

Abbildung 43-5
Diagramm-Editor mit Diagramm



Grundlegende Funktionen des Diagramm-Editors

Im Diagramm-Editor stehen verschiedene Methoden zur Bearbeitung von Diagrammen zur Verfügung.

Menüs

Viele der Aktionen im Diagramm-Editor werden über die Menüs durchgeführt, insbesondere beim Hinzufügen von Objekten zum Diagramm. Beispielsweise werden die Menüs verwendet, um eine Anpassungslinie zu einem Streudiagramm hinzuzufügen. Nach dem Hinzufügen von Objekten zum Diagramm wird häufig das Dialogfeld “Eigenschaften” verwendet, um Optionen für das hinzugefügte Objekt anzugeben.

Dialogfeld “Eigenschaften”

Optionen für das Diagramm und seine Elemente finden Sie im Dialogfeld “Eigenschaften”.

Gehen Sie folgendermaßen vor, um das Dialogfeld “Eigenschaften” aufzurufen:

- Doppelklicken Sie auf ein Diagrammelement.

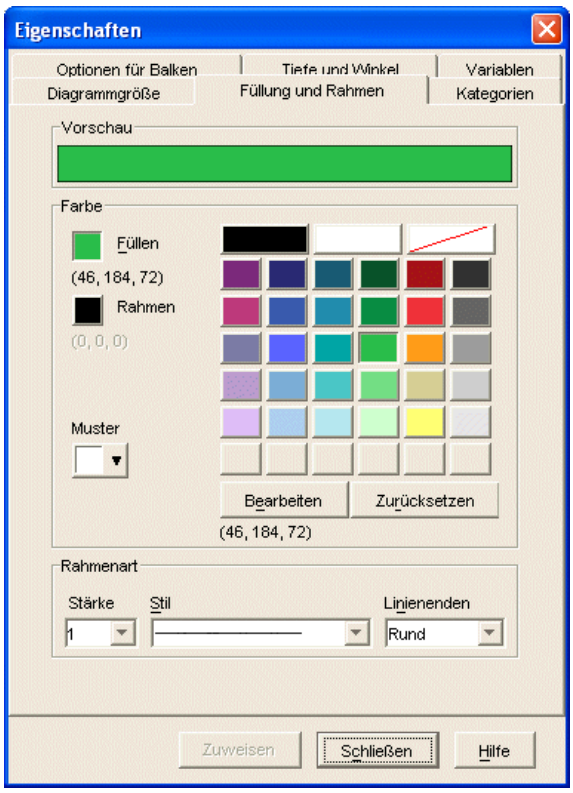
  oder

- Wählen Sie ein Diagrammelement aus und wählen Sie die folgenden Befehle aus den Menüs aus:

  Bearbeiten
  Eigenschaften

Außerdem wird das Dialogfeld “Eigenschaften” automatisch angezeigt, wenn Sie ein Objekt zum Diagramm hinzufügen.

Abbildung 43-6
Dialogfeld “Eigenschaften,” Registerkarte “Füllung und Rahmen”



Das Dialogfeld “Eigenschaften” enthält Registerkarten, mit denen Sie die Optionen festlegen und andere Änderungen an einem Diagramm vornehmen können. Welche Registerkarten im Dialogfeld “Eigenschaften” angezeigt werden, hängt von Ihrer jeweiligen Auswahl ab.

Zu einigen Registerkarten gehört eine Vorschau, die Ihnen eine Vorstellung davon vermittelt, wie sich die Änderungen bei ihrer Anwendung auf die ausgewählten Elemente auswirken. Im Diagramm selbst werden die Änderungen jedoch erst berücksichtigt, nachdem Sie auf Zuweisen geklickt haben. Sie können Änderungen in mehreren Registerkarten vornehmen und erst dann auf Zuweisen klicken. Wenn Sie die Auswahl ändern müssen, um ein anderes Element im Diagramm zu bearbeiten, klicken Sie auf Anwenden, bevor Sie die Auswahl ändern. Wenn Sie nicht vor dem Ändern der Auswahl auf Anwenden klicken, werden die Änderungen beim Klicken auf Anwenden

zu einem späteren Zeitpunkt nur auf das Element oder die Elemente angewendet, die zu diesem Zeitpunkt ausgewählt sind.

Je nach Ihrer Auswahl stehen möglicherweise nur bestimmte Einstellungen zur Verfügung. In der Hilfe für die einzelnen Registerkarten werden die Optionen angegeben, die Sie zur Anzeige der Registerkarten auswählen müssen. Wenn mehrere Elemente ausgewählt wurden, können Sie nur diejenigen Einstellungen ändern, die allen Elementen gemeinsam sind.

Symbolleisten

Über die Symbolleisten können einige der Funktionen im Dialogfeld "Eigenschaften" schneller aufgerufen werden. Beispiel: Anstatt die Registerkarte "Text" im Dialogfeld "Eigenschaften" zu verwenden, können Sie die Schriftart und den Schriftstil eines Texts auch über die Symbolleiste bearbeiten ändern.

Speichern der Änderungen

Änderungen an Diagrammen werden beim Schließen des Diagramm-Editors gespeichert. Das geänderte Diagramm wird anschließend im Viewer angezeigt.

Optionen für die Diagrammdefinition

Wenn Sie ein Diagramm in der Diagrammerstellung definieren, können Sie Titel hinzufügen und die Optionen für die Diagrammerstellung ändern.

Hinzufügen und Bearbeiten von Titeln und Fußnoten

Sie können Titel und Fußnoten zum Diagramm hinzufügen, um seine Interpretation zu erleichtern. Die Diagrammerstellung zeigt außerdem automatisch Fehlerbalkeninformationen in den Fußnoten an.

So können Sie Titel und Fußnoten hinzufügen:

- Klicken Sie auf die Registerkarte Titel/Fußnoten.
- Wählen Sie mindestens einen Titel bzw. eine Fußnote aus. Im Zeichenbereich wird Text angezeigt, um anzugeben, dass diese Elemente zum Diagramm hinzugefügt wurden.
- Im Dialogfeld "Elementeigenschaften" können Sie den Titel-/Fußnotentext bearbeiten.

So entfernen Sie einen Titel bzw. eine Fußnote:

- Klicken Sie auf die Registerkarte Titel/Fußnoten.
- Heben Sie die Auswahl des Titels bzw. der Fußnote, den/die Sie entfernen möchten, auf.

So können Sie den Titel- bzw. Fußnotentext bearbeiten:

Beim Hinzufügen von Titel und Fußnoten kann der zugehörige Text nicht direkt im Diagramm bearbeitet werden. Wie bei anderen Elementen in der Diagrammerstellung nehmen Sie die Bearbeitung im Dialogfeld "Elementeigenschaften" vor.

- Klicken Sie auf Elementeigenschaften, falls das Dialogfeld "Elementeigenschaften" nicht angezeigt wird.
- Wählen Sie in der Liste "Eigenschaften bearbeiten von" einen Titel, einen Untertitel oder eine Fußnote aus (z. B. Titel 1).
- Geben Sie im Inhaltsfeld den Text für den Titel, den Untertitel oder die Fußnote ein.
- Klicken Sie auf Zuweisen.

Festlegen von allgemeinen Optionen

Im Dialogfeld "Diagrammerstellung" sind allgemeine Optionen für das Diagramm verfügbar. Dies sind Optionen, nicht auf ein bestimmtes Element, sondern auf das gesamte Diagramm angewendet werden. Die allgemeinen Optionen schließen die Behandlung von fehlenden Werten, Vorlagen, Diagrammgröße und Feldumbruch ein.

- Klicken Sie auf Optionen.
- Ändern Sie die allgemeinen Optionen. Einzelheiten hierzu finden Sie weiter unten.
- Klicken Sie auf Zuweisen.

Benutzerdefiniert fehlende Werte

Break-Variablen. Wenn bei Variablen, die zum Definieren von Kategorien und Untergruppen verwendet werden, fehlende Werte auftreten, wählen Sie Einschließen, um die Kategorie bzw. Kategorien der benutzerdefinierten fehlenden Werte (Werte, die vom Benutzer als fehlend identifiziert wurden) in das Diagramm einzuschließen. Diese "fehlenden" Kategorien verhalten sich bei der Berechnung der Statistik auch als Break-Variablen. Die Kategorie oder die Kategorien für "Fehlend" werden in der Kategorienachse oder in der Legende angezeigt. Für diese Kategorien werden einem Diagramm beispielsweise zusätzliche Balken oder Kreissegmente hinzugefügt. Wenn keine fehlenden Werte vorhanden sind, werden die Kategorien für "Fehlend" nicht angezeigt.

Wenn Sie diese Option auswählen und die Anzeige nach dem Erstellen des Diagramms unterdrücken möchten, öffnen Sie das Diagramm im Diagramm-Editor und wählen Sie im Menü "Bearbeiten" die Option Eigenschaften. Mit der Registerkarte "Kategorien" können Sie die zu unterdrückenden Kategorien in die Liste "Ausgeschlossen" verschieben. Beachten Sie jedoch, dass die Statistik nicht erneut berechnet wird, wenn Sie die "fehlenden" Kategorien ausblenden. Deswegen werden die "fehlenden" Kategorien bei einer Prozentstatistik beispielsweise immer noch einbezogen.

Anmerkung: Dieses Steuerelement hat keine Auswirkungen auf systemdefiniert fehlende Werte. Diese werden immer aus dem Diagramm ausgeschlossen.

Auswertungsstatistik und Fallwerte. Sie können eine der folgenden Optionen auswählen, um Fälle mit fehlenden Werten auszuschließen:

- **Listenweise ausschließen, um eine konsistente Fallbasis zu erhalten.** Wenn eine der Variablen im Diagramm einen fehlenden Wert für einen bestimmten Fall aufweist, wird der gesamte Fall aus dem Diagramm ausgenommen.
- **Variable für Variable ausschließen, um die Datennutzung zu maximieren.** Wenn eine ausgewählte Variable fehlende Werte aufweist, werden die Fälle mit diesen fehlenden Werten aus der Analyse der Variablen ausgeschlossen.

Der Unterschied zwischen listenweisem und variablenweisem Ausschluss fehlender Werte wird aus den folgenden Abbildungen ersichtlich, die ein Balkendiagramm für jede der beiden Optionen zeigen.

Abbildung 43-7
Listenweiser Ausschluss fehlender Werte

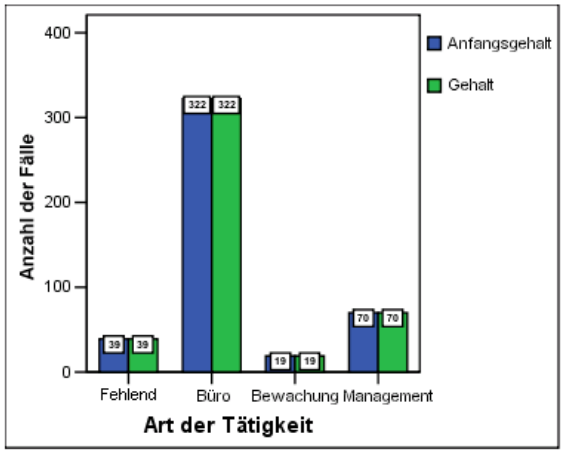
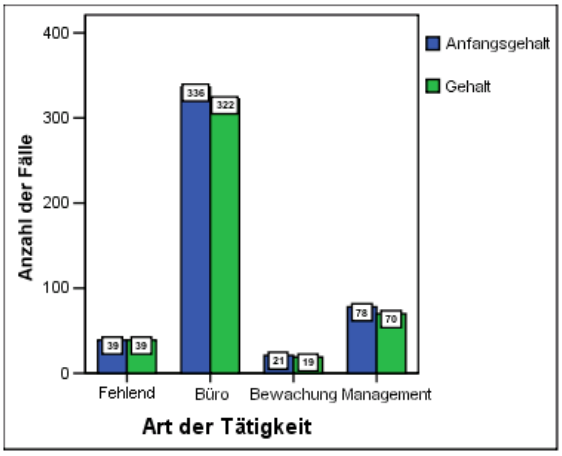


Abbildung 43-8
Ausschluss fehlender Werte Variable für Variable



Die Diagramme wurden aus einer Version der Datei *Employee data.sav* erstellt, die so bearbeitet wurde, dass sie einige systemdefinierte fehlende (leere) Werte in den Variablen für das aktuelle Gehalt und die Art der Tätigkeit aufweist. In anderen Fällen wurde der Wert 0 eingegeben und als

fehlend definiert. Bei beiden Diagrammen wurde die Option Fehlende Werte als Kategorie anzeigen ausgewählt, wodurch die Kategorie *Fehlend* zu den anderen angezeigten Kategorien für die Tätigkeitsart hinzugefügt wird. In jedem Diagramm werden die Werte der Auswertungsfunktion, *Anzahl der Fälle*, in den Balkenbeschriftungen angezeigt.

In beiden Diagrammen weisen 26 Fälle einen systemdefinierten fehlenden Wert für die Art der Tätigkeit auf und 13 Fälle weisen den benutzerdefinierten fehlenden Wert (0) auf. Im Diagramm für den listenweisen Ausschluss ist die Zahl der Fälle für beide Variablen in jeder Balkengruppe gleich, da bei jedem fehlenden Wert der entsprechende Fall für alle Variablen ausgeschlossen wurde. Im Diagramm für den variablenweisen Ausschluss wird die Anzahl der Fälle ohne fehlende Werte für jede Variable in einer Kategorie dargestellt, ohne die fehlenden Werte in anderen Variablen zu berücksichtigen.

Vorlagen zum Definieren von Variablen

Mit einer Diagrammvorlage haben Sie die Möglichkeit, die Attribute eines Diagramms auf ein anderes zu übertragen. Wenn Sie ein Diagramm im Diagramm-Editor öffnen, können Sie dieses als Vorlage speichern. Sie können diese Vorlage dann anwenden, indem Sie sie bei der Erstellung angeben oder später im Diagramm-Editor zuweisen.

Standardvorlage. Hierbei handelt es sich um die Vorlage, die in den Optionen angegeben ist. Sie können auf die Optionen zugreifen, indem Sie im Daten-Editor aus dem Menü “Bearbeiten” den Befehl Optionen auswählen und dann auf die Registerkarte Diagramme klicken. Die Standardvorlage wird zuerst angewendet. Dies bedeutet, dass die anderen Vorlagen diese ersetzen können.

Vorlagendateien. Klicken Sie auf Hinzufügen, um im Standard-Dialogfeld zum Öffnen von Dateien eine oder mehrere Vorlagen anzugeben Diese werden in der Reihenfolge angewendet, in der sie auftreten. Vorlagen am Ende der Liste können also die Vorlagen am Anfang der Liste ersetzen.

Diagrammgröße und Felder

Diagrammgröße. Geben Sie einen Prozentwert über 100 ein, um das Diagramm zu vergrößern, bzw. einen Wert unter 100, um es zu verkleinern. Der Prozentwert bezieht sich auf die Standarddiagrammgröße.

Felder. Wenn viele Feldspalten vorliegen, können Sie mithilfe von Felder umbrechen zulassen, dass die Felder durch Umbrüche auf mehrere Zeilen verteilt werden und nicht in eine Zeile passen müssen. Wenn diese Option nicht ausgewählt ist, werden die Felder soweit verkleinert, bis sie in eine Zeile passen.

Kapitel
44

Extras

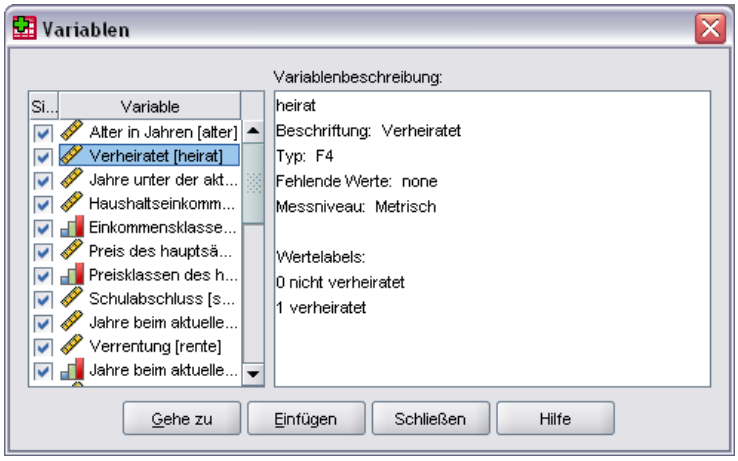
In diesem Kapitel werden die Funktionen aus dem Menü “Extras” und das Umsortieren von Listen mit Zielvariablen behandelt.

Variablenbeschreibungen

Im Dialogfeld “Variablen” werden die folgenden Informationen zur Definition der aktuell ausgewählten Variablen angezeigt:

- Variablenlabel
- Datenformat
- Benutzerdefinierte fehlende Werte
- Wertelabel
- Messniveau

Abbildung 44-1
Dialogfeld “Variablen”



Sichtbar. Die Spalte “Sichtbar” in der Variablenliste zeigt an, ob die Variable aktuell im Dateneditor und in den Variablenlisten der Dialogfelder sichtbar ist. Die Sichtbarkeit wird durch Variablen-Sets festgelegt. Für weitere Informationen siehe Variablen-Sets auf S. 486.

Gehe zu. Hiermit können Sie im Fenster des Daten-Editors zur ausgewählten Variable wechseln.

Einfügen. Hiermit können Sie die ausgewählten Variablen im Haupt-Syntax-Fenster an der Cursorposition einfügen.

Verwenden Sie die Variablenansicht des Daten-Editors zum Ändern der Definitionen von Variablen.

So rufen Sie Variablenbeschreibungen auf:

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Extras
  Variablen
- Wählen Sie die Variable aus, für die Sie Informationen zur Definition aufrufen möchten.

Datendateikommentare

Sie können beschreibende Kommentare in die Datendateien aufnehmen. Bei Datendateien im SPSS-Format werden diese Kommentare zusammen mit den Datendateien gespeichert.

So können Sie Kommentare zu Datendateien hinzufügen, bearbeiten, löschen und anzeigen

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Extras
  Datendateikommentare...
- Um die Kommentare im Viewer anzuzeigen, wählen Sie die Option Kommentare in Ausgabe anzeigen.

Kommentare können beliebig lang sein, sind jedoch auf 80 Byte (bei Single-Byte-Sprachen entspricht dies normalerweise 80 Zeichen) pro Zeile begrenzt; die Zeilen brechen automatisch nach 80 Zeichen um. Kommentare werden in derselben Schriftart angezeigt wie die Textausgabe, um genau widerzuspiegeln, wie sie bei der Anzeige im Viewer dargestellt werden.

Immer, wenn Sie Kommentare hinzufügen oder bearbeiten, wird automatisch ein Datumsstempel (das aktuelle Datum in Klammern) an das Ende der Kommentarliste angehängt. Dies kann zu Unklarheiten hinsichtlich des den Kommentaren zuzuordnenden Datums führen, wenn Sie einen bestehenden Kommentar bearbeiten oder einen neuen Kommentar zwischen bestehenden Kommentaren einfügen.

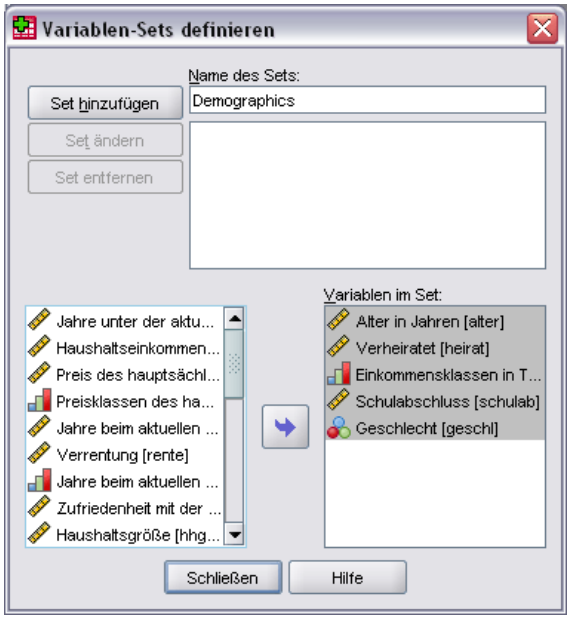
Variablen-Sets

Durch Definieren und Verwenden von Variablen-Sets können Sie einschränken, welche Variablen im Daten-Editor und in den Variablenlisten in den Dialogfeldern angezeigt werden. Dies ist insbesondere bei Datendateien mit einer großen Anzahl an Variablen nützlich. Kleine Variablen-Sets erleichtern das Auffinden und Auswählen von Variablen für die Analysen.

Variablen-Sets definieren

Mit “Variablen-Sets definieren” können Sie Teilmengen von Variablen erstellen, die im Daten-Editor und in den Variablenlisten in den Dialogfeldern angezeigt werden sollen. Die definierten Variablen-Sets werden zusammen mit Datendateien im SPSS-Format gespeichert.

Abbildung 44-2
Dialogfeld "Variablen-Sets definieren"



Name des Sets. Set-Namen können bis zu 64 Byte umfassen. Es können alle Zeichen, einschließlich Leerzeichen, verwendet werden.

Variablen im Set. Ein Set kann eine beliebige Kombination aus numerischen Variablen und String-Variablen enthalten. Die Reihenfolge der Variablen im Set hat keine Auswirkung auf die Anzeigereihenfolge der Variablen im Daten-Editor und in den Variablenlisten in den Dialogfeldern. Eine Variable kann in verschiedenen Sets enthalten sein.

So definieren Sie Variablen-Sets:

- Wählen Sie die folgenden Befehle aus den Menüs aus:
  Extras
  Variablen-Sets definieren...
- Wählen Sie die Variablen aus, die Sie in das Set aufnehmen möchten.
- Geben Sie einen Namen für das Set ein (bis zu 64 Byte).
- Klicken Sie auf Set hinzufügen.

Variablen-Sets verwenden

Mit "Variablen-Sets verwenden" können Sie festlegen, dass im Daten-Editor und in den Variablenlisten in den Dialogfeldern nur die Variablen aus den ausgewählten (markierten) Sets angezeigt werden.

Abbildung 44-3
Dialogfeld "Variablen-Sets verwenden"

- Das im Daten-Editor und in den Variablenlisten in den Dialogfeldern angezeigte Variablen-Set ist die Vereinigungsmenge aller ausgewählten Sets.
- Eine Variable kann in mehreren ausgewählten Sets enthalten sein.
- Die Reihenfolge der Variablen in den ausgewählten Sets und die Reihenfolge der Sets haben keine Auswirkung auf die Anzeigereihenfolge der Variablen im Daten-Editor und in den Variablenlisten in den Dialogfeldern.
- Die definierten Variablen-Sets werden zwar zusammen mit Datendateien im SPSS-Format gespeichert, die Liste der aktuell ausgwählten Sets wird jedoch bei jedem Öffnen der Datendatei auf die standardmäßig integrierten Sets zurückgesetzt.

Die Liste der verfügbaren Variablen-Sets beinhaltet alle Variablen-Sets, die für die Arbeitsdatei definiert wurden, zuzüglich zweier integrierter Sets:

- **ALLVARIABLES.** Dieses Set enthält alle Variablen in der Datendatei sowie die neuen Variablen, die in einer Sitzung erstellt werden.
- **NEWVARIABLES.** Dieses Set enthält nur die in einer Sitzung erstellten Variablen.

  Anmerkung: Selbst wenn Sie die Datendatei nach dem Erstellen neuer Variablen speichern, bleiben diese neuen Variablen weiterhin im Set *NEWVARIABLES* enthalten, bis Sie die Datendatei schließen und erneut öffnen.

Es muss mindestens ein Variablen-Set ausgewählt werden. Wenn *ALLVARIABLES* ausgewählt wurde, haben alle anderen ausgewählten Sets keine sichtbare Wirkung, da dieses Set alle Variablen enthält.

So wählen Sie die anzuzeigenden Variablen-Sets aus:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Extras
  Variablen-Sets verwenden...

▶ Wählen Sie die Variablen-Sets aus, die die Variablen enthalten, die im Daten-Editor und in den Variablenlisten in den Dialogfeldern angezeigt werden sollen.

So zeigen Sie alle Variablen an:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Extras
  Alle Variablen anzeigen

Umsortieren von Listen mit Zielvariablen

Die Variablen in den Listen der Zielvariablen in Dialogfeldern werden in der Reihenfolge angezeigt, in der diese aus der Liste der Quellvariablen ausgewählt wurden. Wenn Sie die Reihenfolge der Variablen in der Liste der Zielvariablen ändern, aber nicht die Auswahl aller Variablen aufheben und dann alle Variablen erneut und in anderer Reihenfolge auswählen möchten, können Sie die Variablen mit der Strg-Taste (Macintosh: Befehlstaste) in der Liste der Zielvariablen nach oben oder nach unten verschieben. Sie können mehrere Variablen auf einmal verschieben, wenn die Variablen in der Liste direkt aufeinander folgen. Sie können keine Gruppen von Variablen verschieben, wenn die Variablen in der Liste nicht direkt aufeinander folgen.

Kapitel
45

Optionen

Im Dialogfeld “Optionen” können Sie eine Vielzahl von Einstellungen ändern, darunter:

- Das Sitzungs-Journal, in dem alle in einer Sitzung verwendeten Befehle aufgezeichnet werden
- Die Reihenfolge der Anzeige von Variablen in den Quelllisten von Dialogfeldern
- Angezeigte und ausgeblendete Objekte in neu ausgegebenen Ergebnissen
- Tabellenvorlage für neue Pivot-Tabellen
- Währungsformate

So ändern Sie die Einstellungen für die Optionen:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Optionen...

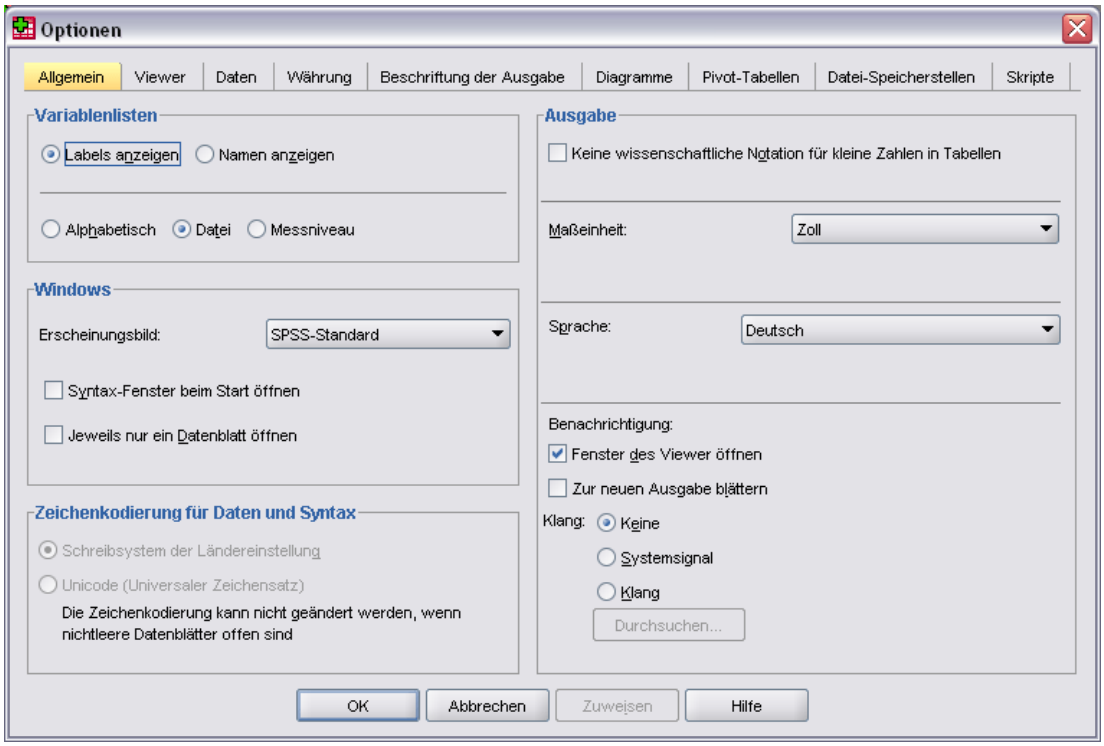
▶ Klicken Sie zum Ändern der Einstellungen auf die entsprechende Registerkarte.

▶ Ändern Sie die Einstellungen.

▶ Klicken Sie auf OK oder Zuweisen.

Optionen: Allgemein

Abbildung 45-1
Dialogfeld "Optionen": Registerkarte "Allgemein"



Variablenlisten

Mit diesen Einstellungen wird festgelegt, wie Variablen in den Listen der Dialogfelder angezeigt werden sollen. Sie können Variablennamen oder Variablenlabels anzeigen lassen. Namen bzw. Labels können in alphabetischer Reihenfolge, in Dateireihenfolge oder nach Messniveau gruppiert angezeigt werden. Die Reihenfolge der Anzeige wirkt sich nur auf Listen von Quellvariablen aus. Listen von Zielvariablen geben immer die Reihenfolge wieder, in welcher die Variablen ausgewählt wurden.

Fenster

Erscheinungsbild. Dient zur Festlegung des grundlegenden Erscheinungsbilds von Fenstern und Dialogfeldern.

Syntax-Fenster beim Start öffnen. Syntax-Fenster sind Fenster für Textdateien zum Eingeben, Bearbeiten und Ausführen von Befehlen. Wenn Sie häufig mit der Befehlssyntax arbeiten, können Sie mit dieser Option zu Beginn jeder SPSS-Sitzung automatisch ein Syntax-Fenster öffnen. Dies bietet sich in erster Linie für erfahrene Anwender an, die lieber mit der Befehlssprache als mit Dialogfeldern arbeiten.

Jeweils nur ein Daten-Set öffnen. Schließt die aktuell geöffnete Datenquelle jedes Mal, wenn über die Menüs und Dialogfelder eine andere Datenquelle geöffnet wird. Standardmäßig gilt: Jedes Mal, wenn Sie über die Menüs und Dialogfelder eine andere Datenquelle öffnen, wird die betreffende Datenquelle in einem neuen Daten-Editor-Fenster angezeigt und alle anderen Datenquellen, die in anderen Fenstern des Daten-Editors geöffnet sind, bleiben während der Sitzung solange geöffnet und verfügbar, bis sie explizit geschlossen werden.

Diese Option wird nach ihrer Auswahl sofort wirksam, schließt jedoch keine Daten-Sets, die zu dem Zeitpunkt geöffnet waren, als die Einstellung geändert wurde. Diese Einstellung hat keine Auswirkungen auf Datenquellen, die mithilfe von Befehlssyntax geöffnet wurden, die auf DATASET-Befehlen zur Steuerung mehrerer Daten-Sets beruht. Für weitere Informationen siehe Arbeiten mit mehreren Datenquellen in Kapitel 6 auf S. 105.

Zeichenkodierung für Daten- und Syntaxdateien

Dadurch wird das Standardverhalten zur Festlegung der Kodierung zum Lesen und Schreiben von Datendateien und Syntaxdateien gesteuert. Diese Einstellung kann nur geändert werden, wenn keine Datenquellen geöffnet sind, und die Einstellung bleibt während der nachfolgenden Sitzungen in Kraft, bis sie explizit geändert wird.

- **Schreibsystem der Ländereinstellung.** Verwendet die Einstellung des aktuellen Gebietsschemas (Ländereinstellung) zum Lesen und Schreiben von Dateien. Diese Vorgehensweise wird auch als **Codepage-Modus** bezeichnet.
- **Unicode (Universaler Zeichensatz).** Unicode-Kodierung (UTF-8) wird zum Lesen und Schreiben von Dateien verwendet. Diese Vorgehensweise wird auch als **Unicode-Modus** bezeichnet.

Es gibt eine Reihe von wichtigen Auswirkungen bei Unicode-Modus und Unicode-Dateien:

- Mit Unicode-Kodierung gespeicherte Datendateien im SPSS-Format sollten nicht in SPSS-Versionen vor 16.0 verwendet werden. Bei Syntaxdateien können Sie beim Speichern der Datei die gewünschte Kodierung angeben. Bei Datendateien müssen Sie die Datendatei im Codepage-Modus öffnen und anschließend erneut speichern, wenn die Datei mit früheren Versionen gelesen werden soll.
- Wenn Codepage-Datendateien im Unicode-Modus gelesen werden, wird die definierte Länge aller Stringvariablen verdreifacht. Um als Länge der einzelnen Stringvariablen automatisch den längsten beobachteten Wert für die betreffende Variable festzulegen, wählen Sie im Dialogfeld “Daten öffnen” die Option String-Längen anhand beobachteter Werte minimieren).

Ausgabe

Keine wissenschaftliche Notation für kleine Zahlen in Tabellen. Hiermit wird die wissenschaftliche Notation bei kleinen Dezimalwerten in der Ausgabe unterdrückt. Sehr kleine Dezimalwerte werden als 0 (oder 0,000) angezeigt.

Maßeinheit. Die Maßeinheit (Punkt, Zoll oder Zentimeter), in der Druckparameter, wie zum Beispiel die Zellenränder von Pivot-Tabellen, Zellenbreiten und Abstand zwischen Tabellen, angegeben werden.

Sprache. Hiermit können Sie die in der Ausgabe verwendete Sprache festlegen. Gilt nicht für einfache Textausgaben, Es hängt von den installierten Sprachdateien ab, welche Sprachen verfügbar sind.

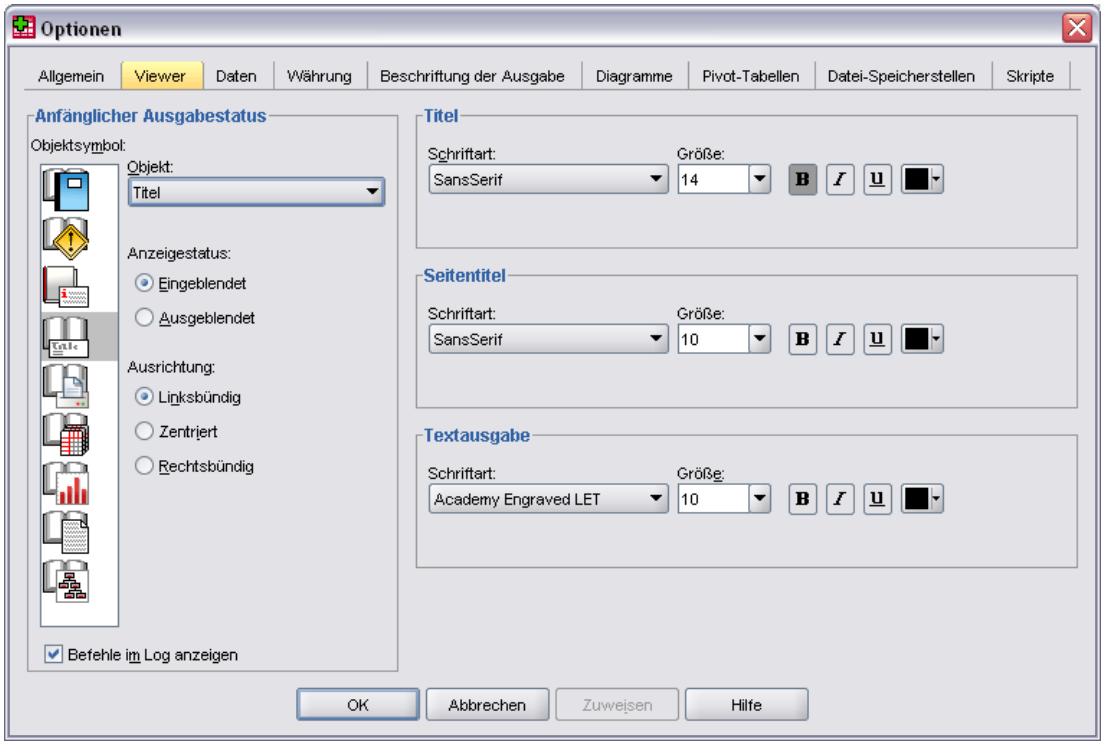
Anmerkung: Benutzerdefinierte Skripte, die sich auf sprachspezifische Zeichenfolgen oder Textpassagen in der Ausgabe stützen, werden möglicherweise nicht ordnungsgemäß ausgeführt, wenn Sie die Ausgabesprache ändern. Für weitere Informationen siehe Optionen: Skripte auf S. 505.

Benachrichtigung. Hiermit können Sie festlegen, wie das Programm Sie über den Abschluss einer Prozedur und die Anzeige der Ergebnisse im Viewer benachrichtigen soll.

Optionen: Viewer

Die Optionen für die Anzeige der Ausgaben wirken sich nur auf die nach dem Ändern der Einstellungen erzeugten Ausgaben aus. Änderungen der Einstellungen wirken sich nicht auf bereits angezeigte Ausgaben aus.

Abbildung 45-2
Dialogfeld "Optionen": Registerkarte "Viewer"



Anfänglicher Ausgabestatus. Hiermit legen Sie fest, welche Objekte bei jeder Ausführung einer Prozedur automatisch angezeigt oder ausgeblendet werden und wie die Objekte ausgerichtet sind. Sie können festlegen, dass folgende Objekte angezeigt bzw. nicht angezeigt werden: Protokoll, Warnungen, Anmerkungen, Titel, Pivot-Tabellen, Diagramme und Textausgabe. Außerdem können Sie die Anzeige von Befehlen im Protokoll aktivieren oder deaktivieren. Sie können Syntaxbefehle aus dem Protokoll kopieren und in einer Syntaxdatei speichern.

Anmerkung: Im Viewer werden alle Ausgabeobjekte linksbündig angezeigt. Die Einstellungen für die Ausrichtung wirken sich nur auf gedruckte Ausgaben aus. Zentrierte und rechtsbündig ausgerichtete Objekte werden durch kleine Symbole an der linken oberen Ecke des Objekts gekennzeichnet.

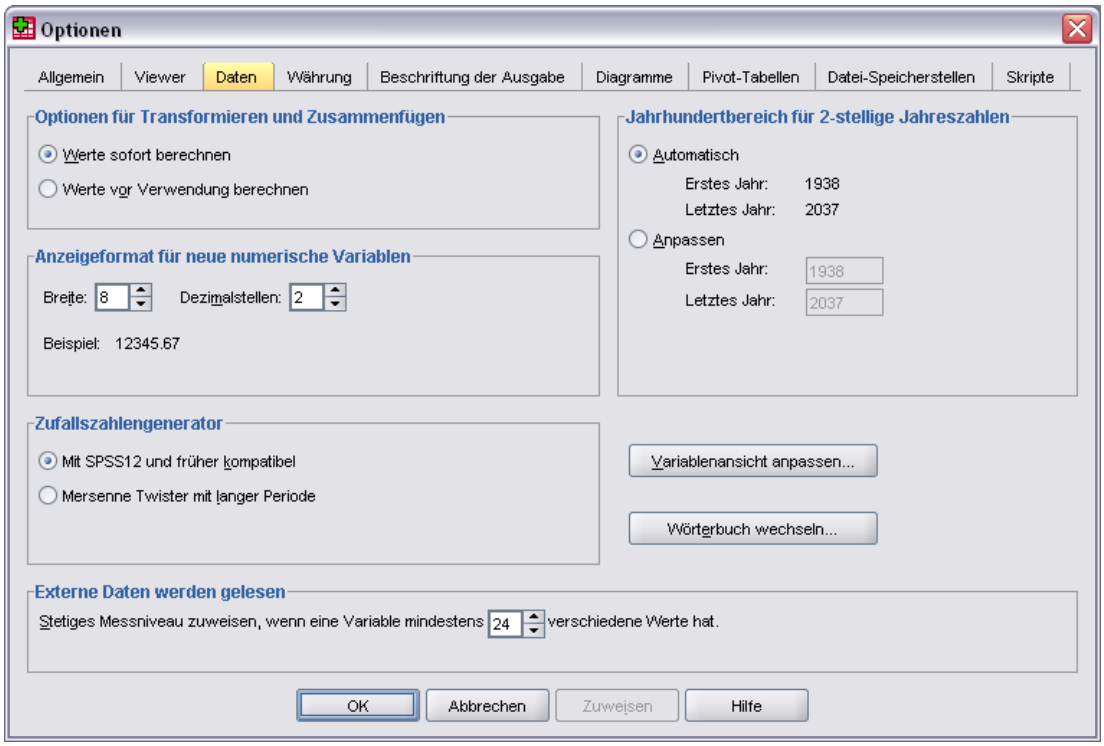
Titel. Hiermit können Sie die Schriftart, Schriftgröße und die Farbe für die Titel neuer Ausgaben festlegen.

Seitentitel. Hiermit legen Sie die Schriftart, Schriftgröße und die Farbe für neue Seitentitel fest, die mit dem Befehl Neuer Seitentitel im Menü "Einfügen" erstellt wurden, und für Seitentitel, die durch die Syntaxbefehl `TITLE` und `SUBTITLE` erzeugt wurden.

Textausgabe. Dies ist die in der Textausgabe verwendete Schriftart. Für die Textausgabe sollte ein Zeichensatz mit festem Abstand verwendet werden. Bei Einsatz von Proportionalschriften werden Tabellenausgaben fehlerhaft ausgerichtet.

Optionen: Daten

Abbildung 45-3
Dialogfeld "Optionen": Registerkarte "Daten"



Optionen für Transformieren und Zusammenfügen. Bei jeder Ausführung eines Befehls wird die Datendatei gelesen. Für einige Datentransformationen, wie beispielsweise "Berechnen" und "Umkodieren", sowie Dateitransformationen, wie beispielsweise "Variablen hinzufügen" und "Fälle hinzufügen", ist kein separater Datendurchlauf erforderlich. Die Ausführung dieser Befehle kann verschoben werden, bis das Programm die Daten liest, um einen anderen Befehl auszuführen, beispielsweise eine Statistik- oder Diagrammprozedur.

- Bei großen Datendateien, bei denen das Einlesen der Daten einige Zeit in Anspruch nehmen kann, sollten Sie die Option Werte vor Verwendung berechnen auswählen, um die Ausführung zu verzögern und Verarbeitungszeit einzusparen. Bei Auswahl dieser Option werden die Ergebnisse von Transformationen, die Sie mithilfe von Registerkarten, wie beispielsweise "Variable berechnen", vornehmen, werden nicht sofort im Daten-Editor angezeigt; neue Variablen, die durch Transformationen entstehen, werden ohne Datenwerte angezeigt und die Datenwerte im Daten-Editor können nicht geändert werden, solange noch Transformationen offen sind. Jeder Befehl, mit dem die Daten gelesen werden, beispielsweise Statistik- oder Diagrammprozeduren, führen die offenen Transformationen aus und aktualisieren die im Daten-Editor angezeigten Daten. Alternativ können Sie die Option Offene Transformationen ausführen im Menü "Transformieren" verwenden.
- Wenn Sie bei der Standardeinstellung Werte sofort berechnen Befehlssyntax aus Dialogfeldern einfügen, wird nach jedem Transformationsbefehl der Befehl EXECUTE eingefügt. Für weitere Informationen siehe Mehrere Execute-Befehle in Kapitel 12 auf S. 274.

Anzeigeformat für neue numerische Variablen. Hiermit können Sie die Standardbreite und Anzahl der Dezimalstellen bei der Anzeige neuer numerischer Variablen festlegen. Es gibt kein Standard-Anzeigeformat für neue String-Variablen. Falls ein Wert zu groß für das festgelegte Anzeigeformat ist, werden erst Dezimalstellen gerundet und dann die Werte in wissenschaftliche Notation umgewandelt. Anzeigeformate haben keine Auswirkung auf die internen Datenwerte. So kann der Wert 123456,78 beispielsweise für die Anzeige auf 123457 gerundet werden, für alle Berechnungen wird jedoch der ursprüngliche, ungerundete Wert verwendet.

Jahrhundertbereich für 2-stellige Jahreszahlen. Hiermit wird der Bereich der Jahre für zweistellig eingegebene und/oder angezeigte Variablen im Datumsformat definiert (zum Beispiel 10/28/86, 29-OKT-87). In der automatischen Einstellung umfaßt der Bereich die 69 Jahre vor und die 30 Jahre nach dem aktuellen Jahr (zusammen mit dem aktuellen Jahr ergibt das 100 Jahre). Bei einem benutzerdefinierten Wert wird das letzte Jahr automatisch anhand des Werts für das erste Jahr berechnet.

Zufallszahlengenerator. Zwei verschiedene Zufallszahlengeneratoren stehen zur Verfügung:

- **Version-12-kompatibel.** Der in Version 12 und früheren Versionen verwendete Zufallszahlengenerator. Wenn Sie randomisierte Ergebnisse reproduzieren möchten, die in früheren Versionen auf der Grundlage eines angegebenen Startwerts generiert wurden, müssen Sie diesen Zufallszahlengenerator verwenden.
- **Mersenne-Twister.** Ein neuerer Zufallszahlengenerator, der für Simulationszwecke eine höhere Zuverlässigkeit bietet. Sofern es nicht darum geht, zufallsbestimmte Ergebnisse aus SPSS 12 oder älteren Versionen zu reproduzieren, sollte dieser Zufallszahlengenerator verwendet werden.

Externe Daten werden gelesen. Bei Daten, die aus externen Dateiformaten und Datendateien in Formaten von älteren Versionen von SPSS (vor Version 8.0) erstellt wurden, können Sie für eine numerische Variable die Mindestanzahl an Datenwerten angeben, anhand deren die Variable als metrische oder nominale Variable klassifiziert wird. Variablen mit einer geringeren als der angegebenen Anzahl von eindeutigen Werten werden als nominal klassifiziert.

Variablenansicht anpassen. Dient zur Festlegung der Standardvorgaben für Anzeige und Reihenfolge der Attribute in der Variablenansicht. Für weitere Informationen siehe Ändern der Standard-Variablenansicht auf S. 496.

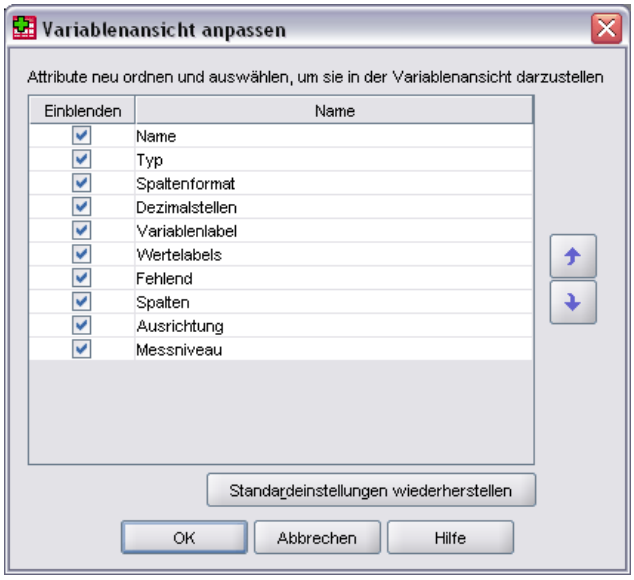
Wörterbuch wechseln. Dient zur Festlegung der Sprachversion des Wörterbuchs, das für die Rechtschreibprüfung der Elemente in der Variablenansicht verwendet wird. Für weitere Informationen siehe Rechtschreibprüfung bei Variablen- und Wertelabels in Kapitel 5 auf S. 96.

Ändern der Standard-Variablenansicht

Mit "Variablenansicht anpassen" können Sie festlegen, welche Attribute standardmäßig in der Variablenansicht angezeigt werden (z. B. Name, Typ, Variablenlabel) und in welcher Reihenfolge diese Attribute angezeigt werden.

Klicken Sie auf Variablenansicht anpassen.

Abbildung 45-4
Variablenansicht anpassen (Standard)



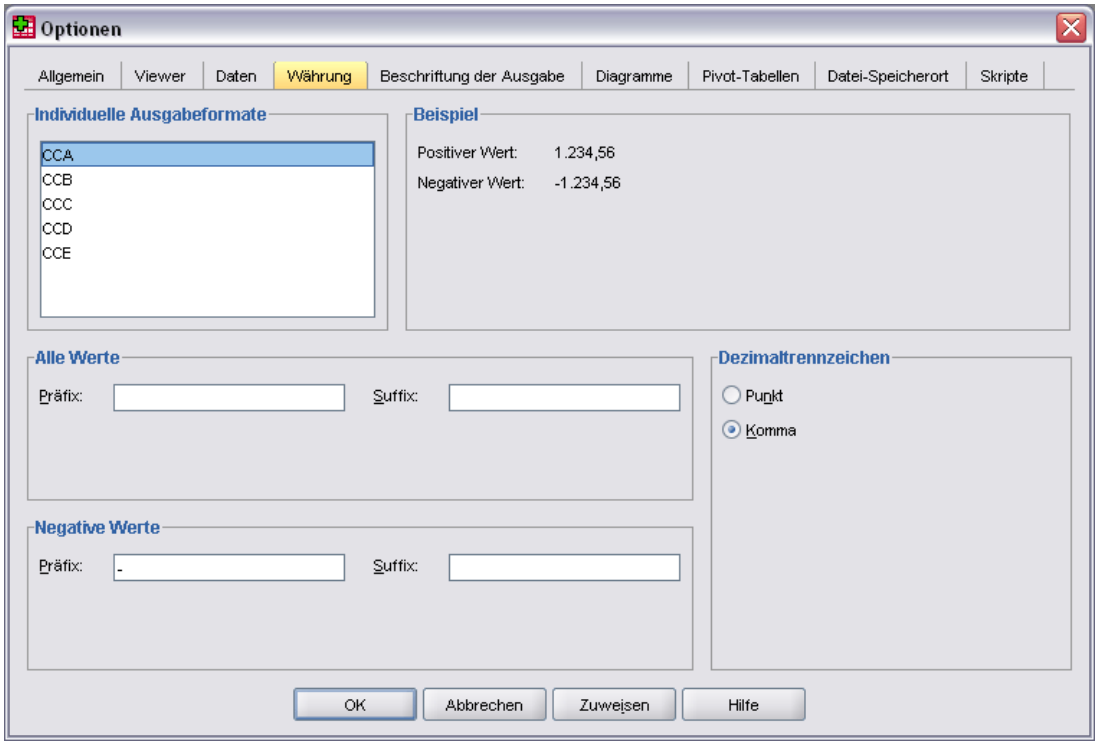
- Aktivieren Sie die Variablenattribute, die angezeigt werden sollen.
- Die Anzeigereihenfolge der Attribute können Sie mit der nach oben bzw. nach unten weisenden Pfeilschaltfläche ändern.

Optionen: Währung

Sie können bis zu fünf spezielle Anzeigeformate für Währungen erstellen, die über jeweils ein spezielles Präfix und Suffix verfügen und eine spezielle Behandlung negativer Werte beinhalten können.

Die fünf Namen der benutzerdefinierten Währungsfomate lauten CCA, CCB, CCC, CCD und CCE. Sie können die Namen der Formate nicht ändern und keine neuen hinzufügen. Sie können ein Währungsformat ändern, indem Sie den Formatnamen aus der Quellliste auswählen und die gewünschten Änderungen vornehmen.

Abbildung 45-5
Dialogfeld "Optionen": Registerkarte "Währung"



Die für die Währungsformate definierten Präfixe, Suffixe und Dezimalzeichen dienen nur zur Anzeige. Sie können im Daten-Editor keine Werte mit Zeichen für spezielle Währungen eingeben.

So erstellen Sie Währungsformate:

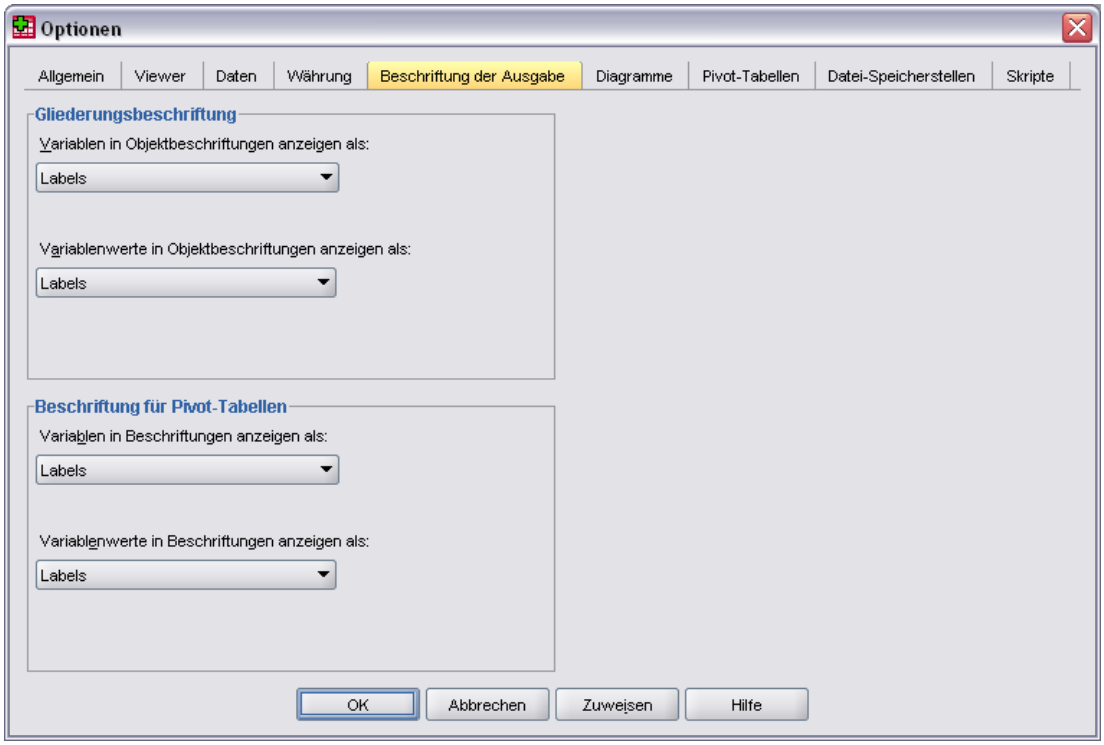
- ▶ Klicken Sie auf die Registerkarte Währung.
- ▶ Wählen Sie eines der Währungsformate (CCA, CCB, CCC, CCD oder CCE) aus der Liste aus.
- ▶ Geben Sie das Präfix, das Suffix und ein Dezimaltrennzeichen ein.
- ▶ Klicken Sie auf OK oder Zuweisen.

Optionen: Beschriftung der Ausgabe

Mit den Optionen für die Beschriftung der Ausgabe können Sie Einstellungen für die Anzeige der Variablen und Datenwerte in der Gliederung und in Pivot-Tabellen vornehmen. Sie können Variablennamen, definierte Variablenlabels und Datenwerte, definierte Wertelabels oder eine Kombination dieser Angaben anzeigen lassen.

Aussagekräftige Variablen- und Wertelabels (Variablenansicht im Daten-Editor, Spalten *Variablenlabel* und *Wertelabels*) können die Interpretation der Ergebnisse häufig erleichtern. In manchen Tabellen können lange Labels jedoch eher stören.

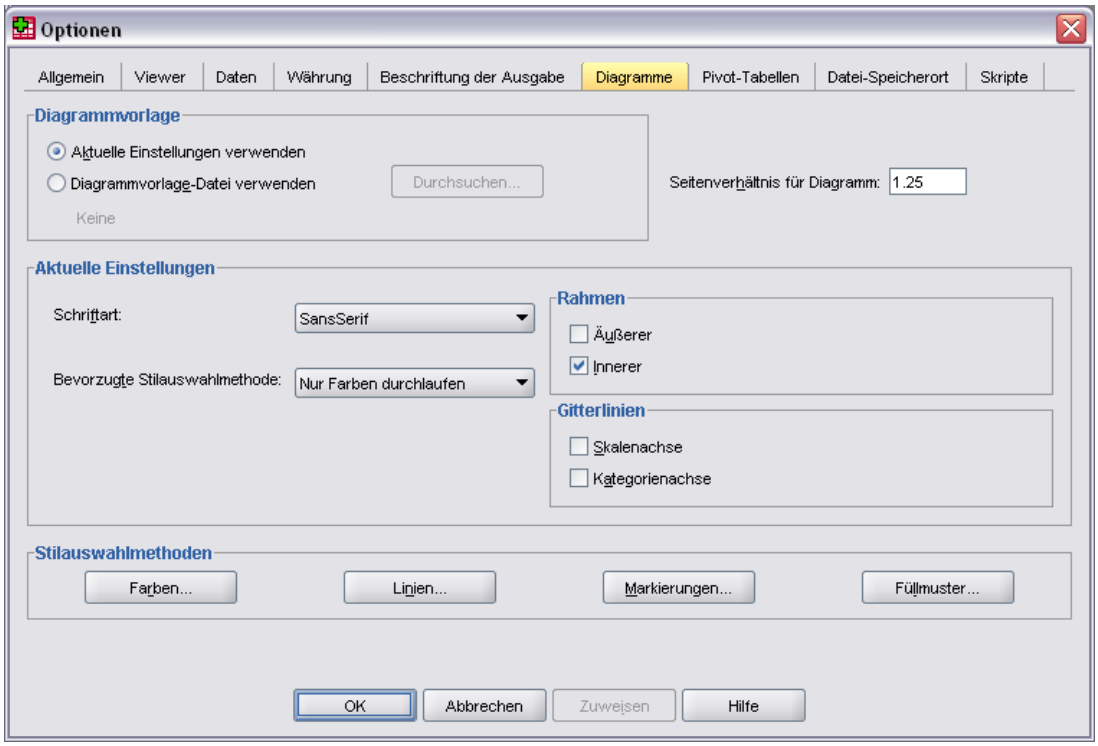
Abbildung 45-6
Dialogfeld "Optionen": Registerkarte "Beschriftung der Ausgabe"



Die Optionen für die Beschriftung der Ausgabe wirken sich nur auf neue Ausgaben aus, die nach der Änderung der Einstellungen berechnet werden. Änderungen der Einstellungen wirken sich nicht auf bereits angezeigte Ausgaben aus. Diese Einstellungen betreffen nur die Ausgabe von Pivot-Tabellen. Sie wirken sich nicht auf die Textausgabe aus.

Diagrammoptionen

Abbildung 45-7
Dialogfeld "Optionen": Registerkarte "Diagramme"



Diagrammvorlage. Sie können für neue Diagramme entweder die hier ausgewählten Einstellungen oder die Einstellungen aus einer Diagrammvorlagendatei auswählen. Klicken Sie auf Durchsuchen, um eine Diagrammvorlagendatei auszuwählen. Sie erstellen eine Diagrammvorlagendatei, indem Sie eine Datei mit den von Ihnen gewünschten Merkmalen erstellen und im Menü "Datei" den Befehl Diagrammvorlage speichern auswählen.

Seitenverhältnis für Diagramm. Dies ist das Verhältnis von Breite zu Höhe des äußeren Rahmens neuer Diagramme. Sie können ein Verhältnis von Breite zu Höhe von 0,1 bis 10,0 festlegen. Werte unter 1 ergeben Diagramme im Hochformat. Werte größer als 1 ergeben Diagramme im Querformat. Ein Wert von 1 ergibt quadratische Diagramme. Bei bereits erstellten Diagrammen kann das Seitenverhältnis nicht mehr geändert werden.

Aktuelle Einstellungen. Folgende Einstellungen sind verfügbar:

- **Schriftart.** Diese Schriftart wird für den gesamten Text in neuen Diagrammen verwendet.
- **Bevorzugte Stilauswahlmethode.** Hiermit geben Sie die Anfangswerte für Farben und/oder Muster für neue Diagramme ein. Bei Nur Farben durchlaufen werden zum Unterscheiden verschiedener Diagrammelemente keine Muster, sondern nur Farben verwendet. Bei Nur Muster durchlaufen werden zum Unterscheiden verschiedener Diagrammelemente keine Farben, sondern nur Linienstile, Markierungssymbole und Füllmuster verwendet.

- **Rahmen.** Hiermit können Sie festlegen, ob neue Diagramme mit inneren bzw. äußeren Rahmen erstellt werden sollen.
- **Gitterlinien.** Hiermit können Sie festlegen, ob neue Diagramme mit Gitterlinien für die Skalen- und Kategorienachse angezeigt werden sollen.

Stilauswahlmethoden. Dient zur benutzerdefinierten Anpassung der Farben, Linienstile, Markierungssymbole und Füllmuster für neue Diagramme. Sie können die Anordnung der Farben und Muster ändern, die beim Erstellen eines neuen Diagramms verwendet werden.

Datenelement Farben

Geben Sie die Reihenfolge an, in der die Farben im neuen Diagramm für die Datenelemente (z. B. Balken und Markierungen) verwendet werden sollen. Farben werden immer dann verwendet, wenn Sie eine Auswahl treffen, zu der im Hauptdialogfeld "Diagrammoptionen" in der Gruppe "Bevorzugte Stilauswahlmethode" die Option *Farbe* gehört.

Wenn Sie beispielsweise ein gruppiertes Balkendiagramm mit zwei Gruppen erstellen und im Hauptdialogfeld "Diagrammoptionen" die Option Erst Farbpalette, dann Muster durchlaufen auswählen, werden die ersten beiden Farben in der Liste der gruppierten Diagramme im neuen Diagramm als Balkenfarben verwendet.

So ändern Sie die Reihenfolge, in der die Farben verwendet werden:

▶ Wählen Sie die Option Einfache Diagramme und wählen Sie dann eine Farbe aus, die für Diagramme ohne Kategorien verwendet werden soll.

▶ Wählen Sie die Option Gruppierte Diagramme, um die Farbauswahlmethode für Diagramme mit Kategorien zu ändern. Wenn Sie die Farbe einer Kategorie ändern möchten, wählen Sie die Kategorie und anschließend eine Farbe für diese Kategorie aus der Farbpalette aus.

Die folgenden Optionen sind verfügbar:

- Sie können vor der jeweils ausgewählten Kategorie eine neue Kategorie einfügen.
- Ausgewählte Kategorien können verschoben werden.
- Ausgewählte Kategorien können entfernt werden.
- Die Sequenz kann auf den Standardwert zurückgesetzt werden.
- Sie können eine Farbe bearbeiten, indem Sie ihre Quelle auswählen und auf Bearbeiten klicken.

Linien von Datenelementen

Geben Sie die Reihenfolge an, in der die verschiedenen Stile für linienförmige Datenelemente in Ihrem neuen Diagramm verwendet werden sollen. Linienstile werden immer dann verwendet, wenn das Diagramm linienförmige Datenelemente enthält und Sie eine Auswahl treffen, zu der im Hauptdialogfeld "Diagrammoptionen" in der Gruppe "Bevorzugte Stilauswahlmethode" die Option *Muster* gehört.

Wenn Sie beispielsweise ein Liniendiagramm mit zwei Gruppen erstellen und im Hauptdialogfeld "Diagrammoptionen" die Option Nur Muster durchlaufen auswählen, werden die ersten beiden Stile in der Liste der gruppierten Diagramme im neuen Diagramm als Linienmuster verwendet.

So ändern Sie die Reihenfolge, in der die Linienmuster verwendet werden:

- Wählen Sie die Option Einfache Diagramme und wählen Sie dann einen Linienstil aus, der für Liniendiagramme ohne Kategorien verwendet werden soll.
- Wählen Sie die Option Gruppierte Diagramme, um die Musterauswahlmethode für Diagramme mit Kategorien zu ändern. Wenn Sie den Linienstil einer Kategorie ändern möchten, wählen Sie die Kategorie und anschließend einen Linienstil für diese Kategorie aus der Palette aus.

Die folgenden Optionen sind verfügbar:

- Sie können vor der jeweils ausgewählten Kategorie eine neue Kategorie einfügen.
- Ausgewählte Kategorien können verschoben werden.
- Ausgewählte Kategorien können entfernt werden.
- Die Sequenz kann auf den Standardwert zurückgesetzt werden.

Markierungen für Datenelemente

Geben Sie die Reihenfolge an, in der die verschiedenen Symbole für Markierungs-Datenelemente in Ihrem neuen Diagramm verwendet werden sollen. Markierungsstile werden immer dann verwendet, wenn das Diagramm Markierungs-Datenelemente enthält und Sie eine Auswahl treffen, zu der im Hauptdialogfeld "Diagrammoptionen" in der Gruppe "Bevorzugte Stilauswahlmethode" die Option *Muster* gehört.

Wenn Sie beispielsweise ein Streudiagramm mit zwei Gruppen erstellen und im Hauptdialogfeld "Diagrammoptionen" die Option Nur Muster durchlaufen auswählen, werden die ersten beiden Symbole in der Liste der gruppierten Diagramme im neuen Diagramm als Markierungen verwendet.

So ändern Sie die Reihenfolge, in der die Markierungsstile verwendet werden:

- Wählen Sie die Option Einfache Diagramme und wählen Sie dann ein Symbol aus, das für Diagramme ohne Kategorien verwendet werden soll.
- Wählen Sie die Option Gruppierte Diagramme, um die Musterauswahlmethode für Diagramme mit Kategorien zu ändern. Wenn Sie das Markierungssymbol einer Kategorie ändern möchten, wählen Sie die Kategorie und anschließend ein Symbol für diese Kategorie aus der Palette aus.

Die folgenden Optionen sind verfügbar:

- Sie können vor der jeweils ausgewählten Kategorie eine neue Kategorie einfügen.
- Ausgewählte Kategorien können verschoben werden.
- Ausgewählte Kategorien können entfernt werden.
- Die Sequenz kann auf den Standardwert zurückgesetzt werden.

Füllmuster für Datenelemente

Geben Sie die Reihenfolge an, in der die verschiedenen Füllstile für Balken- und Flächen-Datenelemente in Ihrem neuen Diagramm verwendet werden sollen. Füllstile werden immer dann verwendet, wenn das Diagramm Balken- oder Flächen-Datenelemente enthält und Sie eine Auswahl treffen, zu der im Hauptdialogfeld "Diagrammoptionen" in der Gruppe "Bevorzugte Stilauswahlmethode" die Option *Muster* gehört.

Wenn Sie beispielsweise ein gruppiertes Balkendiagramm mit zwei Gruppen erstellen und im Hauptdialogfeld "Diagrammoptionen" die Option Nur Muster durchlaufen auswählen, werden die ersten beiden Stile in der Liste der gruppierten Diagramme im neuen Diagramm als Füllmuster für die Balken verwendet.

So ändern Sie die Reihenfolge, in der die Füllstile verwendet werden:

- Wählen Sie die Option Einfache Diagramme und wählen Sie dann ein Füllmuster aus, das für Diagramme ohne Kategorien verwendet werden soll.
- Wählen Sie die Option Gruppierte Diagramme, um die Musterauswahlmethode für Diagramme mit Kategorien zu ändern. Wenn Sie das Füllmuster einer Kategorie ändern möchten, wählen Sie die Kategorie und anschließend ein Füllmuster für diese Kategorie aus der Palette aus.

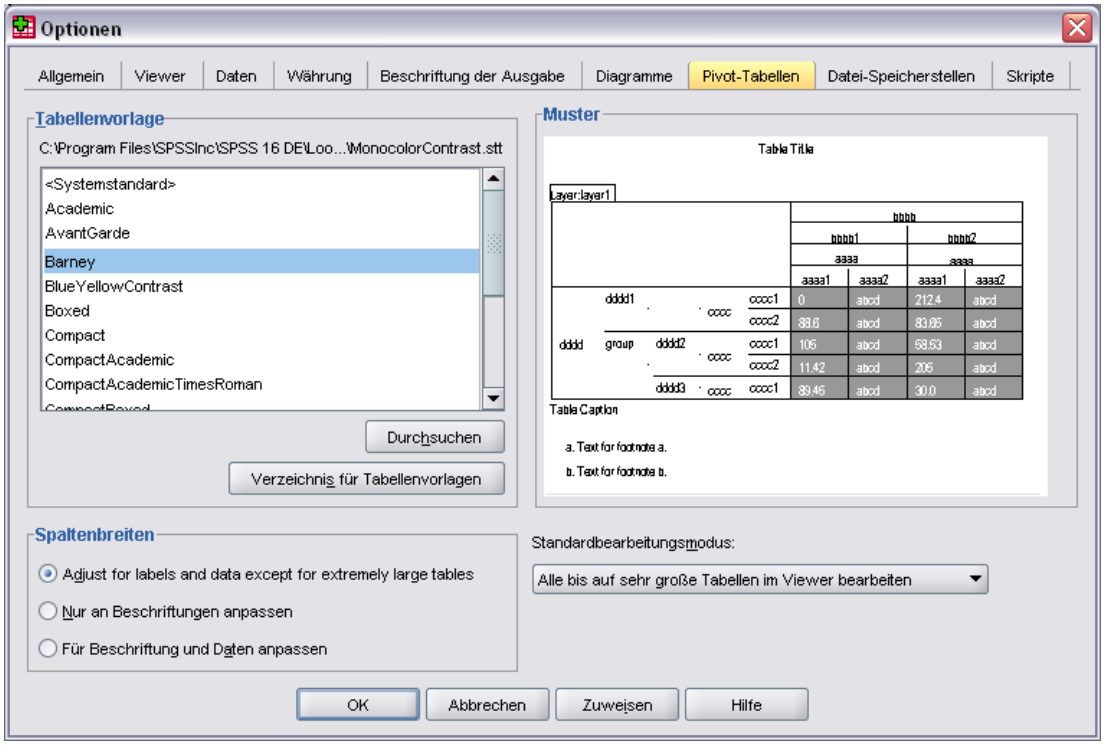
Die folgenden Optionen sind verfügbar:

- Sie können vor der jeweils ausgewählten Kategorie eine neue Kategorie einfügen.
- Ausgewählte Kategorien können verschoben werden.
- Ausgewählte Kategorien können entfernt werden.
- Die Sequenz kann auf den Standardwert zurückgesetzt werden.

Pivottabellenoptionen

Mit den Optionen für Pivot-Tabellen können Sie die Standard-Tabellenvorlage einstellen, die für neue Pivot-Tabellen verwendet werden soll. Mit den Tabellenvorlagen können Sie eine Reihe von Parametern für Pivot-Tabellen einstellen, darunter die Anzeige und Breite von Gitterlinien, Schriftart, Schriftgröße und -farbe sowie Hintergrundfarben.

Abbildung 45-8
Dialogfeld "Optionen": Registerkarte "Pivot-Tabellen"



Tabellenvorlage. Wählen Sie eine Tabellenvorlage aus der Dateiliste aus und klicken Sie auf OK oder Übernehmen. Sie können eine der mit SPSS gelieferten Tabellenvorlagen verwenden oder im Pivot-Tabellen-Editor Ihre eigene Tabellenvorlage erstellen. Wählen Sie dazu im Menü "Format" die Option Tabellenvorlagen.

- **Durchsuchen.** Hiermit können Sie eine Tabellenvorlage aus einem anderen Verzeichnis auswählen.
- **Verzeichnis für Tabellenvorlagen.** Hiermit können Sie ein anderes Standardverzeichnis für Tabellenvorlagen angeben.

Anmerkung: Tabellenvorlagen, die in früheren Versionen von SPSS erstellt wurden, können nicht in Version 16.0 oder später verwendet werden.

Spaltenbreite einstellen für. Hiermit können Sie die automatische Anpassung der Spaltenbreite in Pivot-Tabellen einstellen.

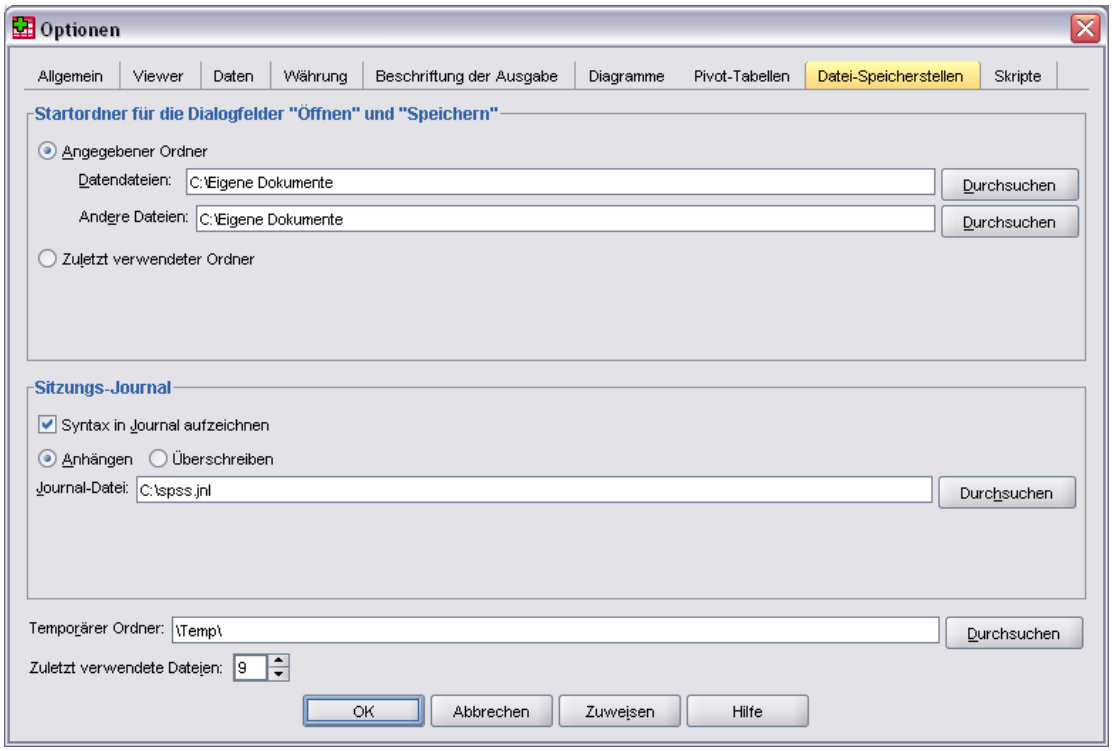
- **Beschriftungen.** Hiermit passen Sie die Spaltenbreite an die Breite der Spaltenbeschriftung an. Dadurch erhalten Sie kompaktere Tabellen, aber Datenwerte, die breiter als die Beschriftung sind, werden nicht angezeigt. (Sternchen weisen auf Werte hin, die zu breit für die Anzeige sind.)
- **Beschriftungen und Daten.** Hiermit passen Sie die Spaltenbreite an die Spaltenbeschriftung oder den größten Datenwert an, je nachdem, welcher der beiden länger ist. Dies erzeugt breitere Tabellen, stellt jedoch die Anzeige aller Werte sicher.

Standardbearbeitungsmodus. Hiermit können Sie einstellen, ob Pivot-Tabellen im Viewer-Fenster oder in einem separaten Fenster aktiviert werden. In der Standardeinstellung werden durch Doppelklicken auf eine Pivot-Tabelle alle Tabellen (bis auf extrem große Tabellen) im Viewer-Fenster aktiviert. Sie können Pivot-Tabellen jedoch auch in einem separaten Fenster öffnen oder festlegen, dass kleine Pivot-Tabellen im Viewer-Fenster und Pivot-Tabellen ab einer bestimmten Größe in einem separaten Fenster geöffnet werden.

Optionen für Datei-Speicherstellen

Mit den Optionen auf der Registerkarte "Datei-Speicherort" können Sie den Standard-Speicherort festlegen, den die Anwendung zu Beginn jeder Sitzung zum Öffnen und Speichern von Dateien verwendet, den Speicherort der "Journal-Datei", den Speicherort des temporären Ordners und die Anzahl der Dateien, die in der Liste der zuletzt verwendeten Dateien angezeigt werden.

Abbildung 45-9
Dialogfeld "Optionen": Registerkarte "Datei-Speicherort"



Startordner für die Dialogfelder "Öffnen" und "Speichern"

- **Angegebener Ordner.** Der angegebene Ordner dient als Standard-Speicherort zu Beginn jeder Sitzung. Sie können unterschiedliche Standard-Speicherstellen für Datendateien und sonstige Dateien angeben.
- **Zuletzt verwendeter Ordner.** Der in der vorangegangenen Sitzung zuletzt zum Öffnen bzw. Speichern von Dateien verwendete Ordner wird standardmäßig beim Start der nächsten Sitzung verwendet. Dies gilt sowohl für Datendateien als auch für sonstige Dateien.

Diese Einstellungen gelten nur für Dialogfelder zum Öffnen und Speichern von Dateien und der "zuletzt verwendete Ordner" wird aus dem letzten Dialogfeld ermittelt, das zum Öffnen bzw. Speichern einer Datei verwendet wurde. Dateien, die über Befehlssyntax geöffnet bzw. gespeichert wurden, haben keine Wirkung auf diese Einstellungen und sind auch nicht von ihnen betroffen.

Sitzungs-Journal

Mit dem Sitzungs-Journal können Sie automatisch die in einer Sitzung ausgeführten Befehle aufzeichnen. Dazu gehören Befehle, die in Syntax-Fenstern eingegeben und von dort ausgeführt wurden, und aus Dialogfeldern aufgerufene Befehle. Sie können die Journaldatei bearbeiten und die Befehle erneut in anderen Sitzungen verwenden. Sie können das Führen der Journaldatei aktivieren und deaktivieren, etwas an die Journaldatei anhängen oder die Datei überschreiben sowie Namen und Speicherort der Journaldatei auswählen. Sie können Syntaxbefehle aus der Journaldatei kopieren und in einer Syntaxdatei speichern.

Temporärer Ordner

Dient zur Festlegung des Speicherorts für die temporären Dateien, die in einer Sitzung erstellt werden. Der Speicherort für temporäre Datendateien im Modus für verteilte Analysen (verfügbar mit der Server-Version) wird hierdurch nicht beeinflusst. Im Modus für verteilte Analysen wird der Speicherort für temporäre Dateien durch die Umgebungsvariable *SPSSTMPDIR* festgelegt. Diese Variable kann nur auf dem Computer gesetzt werden, auf dem die Server-Version der Software ausgeführt wird. Wenn Sie den Speicherort für temporäre Dateien ändern möchten, wenden Sie sich an Ihren Systemadministrator.

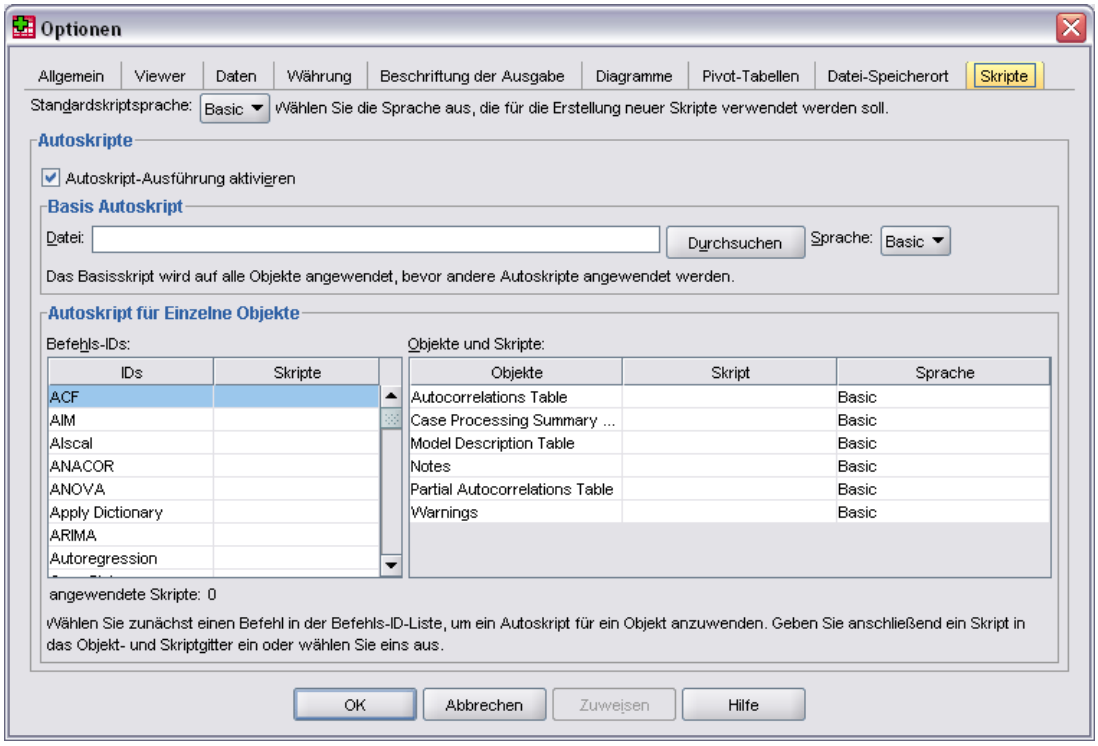
Zuletzt verwendete Dateien

Hiermit legen Sie die Anzahl der im Menü "Datei" angezeigten zuletzt verwendeten Dateien fest.

Optionen: Skripte

Auf der Registerkarte "Skripte" können Sie die Standard-Skriptsprache und alle etwaigen Autoskripts angeben, die Sie verwenden möchten. Sie können Skripte zum Automatisieren vieler Funktionen verwenden, beispielsweise zum Anpassen von Pivot-Tabellen.

Abbildung 45-10
Dialogfeld "Optionen": Registerkarte "Skripte"



Anmerkung: Benutzer, die veraltete Sax Basic-Sprache verwenden, müssen etwaige benutzerdefinierte Autoskripts manuell konvertieren. Die mit Versionen vor 16.0 installieren Autoskripts sind als Set separater Skriptdateien verfügbar und befinden sich im Unterverzeichnis *Samples* des Installationsverzeichnisses von SPSS. Standardmäßig sind Autoskripts keinerlei Ausgabeobjekte zugeordnet. Sie müssen alle Autoskripts manuell mit den gewünschten Ausgabeobjekten verknüpfen, wie unten beschrieben.

Standard-Skriptsprache. Die Standard-Skriptsprache bestimmt den Skript-Editor, der gestartet wird, wenn neue Skripte erstellt werden. Außerdem gibt sie die Standardsprache an, deren ausführbares Element zur Ausführung von Autoskripts verwendet wird. Welche Skriptsprachen verfügbar sind, hängt von der verwendeten Plattform ab. Unter Windows sind als Skriptsprachen die Sprache "Basic", die zusammen mit dem Base-System installiert wird, und die Programmiersprache "Python" verfügbar. Bei allen anderen Plattformen steht für Skripte die Programmiersprache Python zur Verfügung.

Anmerkung:

Autoskript-Ausführung aktivieren. Mit diesem Kontrollkästchen können Sie die Autoskript-Ausführung aktivieren bzw. deaktivieren. Standardmäßig ist die Autoskript-Ausführung aktiviert.

Basis Autoskript. Ein optionales Skript, das vor allen anderen Autoskripts auf alle neuen Viewer-Objekte angewendet wird. Geben Sie die Skriptdatei an, die als Basis-Autoskript verwendet werden soll, sowie die Sprache deren ausführbares Element zur Ausführung des Skripte verwendet werden sollen.

So wenden Sie Autoskripts auf Ausgabeobjekte an:

- Wählen Sie im Befehls-ID-Gitter einen Befehl aus, der Ausgabeobjekte generiert, auf die Autoskripts angewendet werden.

  In der Spalte *Objekte* im Gitter "Objekte und Skripte" wird eine Liste der Objekte angezeigt, die mit dem ausgewählten Befehl verknüpft sind. In der Spalte *Skript* werden etwaige bestehende Skripte für den ausgewählten Befehl angezeigt.

- Geben Sie ein Skript für alle Elemente an, die in der Spalte *Objekte* angezeigt werden. Klicken Sie auf die entsprechende *Skript*-Zelle. Geben Sie den Pfad zu dem Skript ein oder klicken Sie auf die Schaltfläche mit den Auslassungszeichen (...), um nach dem Skript zu suchen.

- Geben Sie die Sprache an, deren ausführbares Element zur Skriptausführung verwendet werden soll. *Anmerkung*: Eine Änderung der Standard-Skriptsprache hat keine Auswirkungen auf die ausgewählte Sprache.

- Klicken Sie auf Zuweisen oder OK.

So entfernen Sie Autoskript-Zuordnungen:

- Klicken Sie im Gitter "Objekte und Skripte" auf die Zelle in der Spalte "Skript", die zu dem Skript gehört, dessen Zuordnung Sie aufheben möchten.

- Löschen Sie den Pfad zu dem Skript und klicken Sie dann auf eine andere Zelle im Gitter "Objekte und Skripte".

- Klicken Sie auf Zuweisen oder OK.

So legen Sie Skriptoptionen fest:

Wählen Sie die folgenden Befehle aus den Menüs aus:

Bearbeiten
  Optionen

- Klicken Sie auf die Registerkarte "Skripte".

- Wählen Sie die gewünschten Einstellungen aus.

- Klicken Sie auf Zuweisen oder OK.

Kapitel
46

Anpassen von Menüs und Symbolleisten

Menü-Editor

Sie können den Menü-Editor zum Anpassen der Menüs von SPSS verwenden. Mit dem Menü-Editor stehen Ihnen die folgenden Möglichkeiten zur Verfügung:

- Sie können Einträge zu Menüs hinzufügen, mit denen angepaßte SPSS-Skripts ausgeführt werden.
- Sie können Einträge zu Menüs hinzufügen, mit denen SPSS-Befehlssyntax-Dateien ausgeführt werden.
- Sie können Einträge zu Menüs hinzufügen, mit denen andere Anwendungen gestartet und Daten aus SPSS automatisch an andere Anwendungen übergeben werden.

Sie können Daten in den folgenden Formaten an andere Anwendungen versenden: SPSS, Excel, Lotus 1-2-3, durch Tabulatoren getrennt, und dBASE IV.

So fügen Sie den Menüs von SPSS Einträge hinzu:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Ansicht
  Menü-Editor...

▶ Doppelklicken Sie im Dialogfeld "Menü-Editor" auf das Menü, dem Sie einen neuen Eintrag hinzufügen möchten (oder klicken Sie auf das Symbol mit dem Pluszeichen).

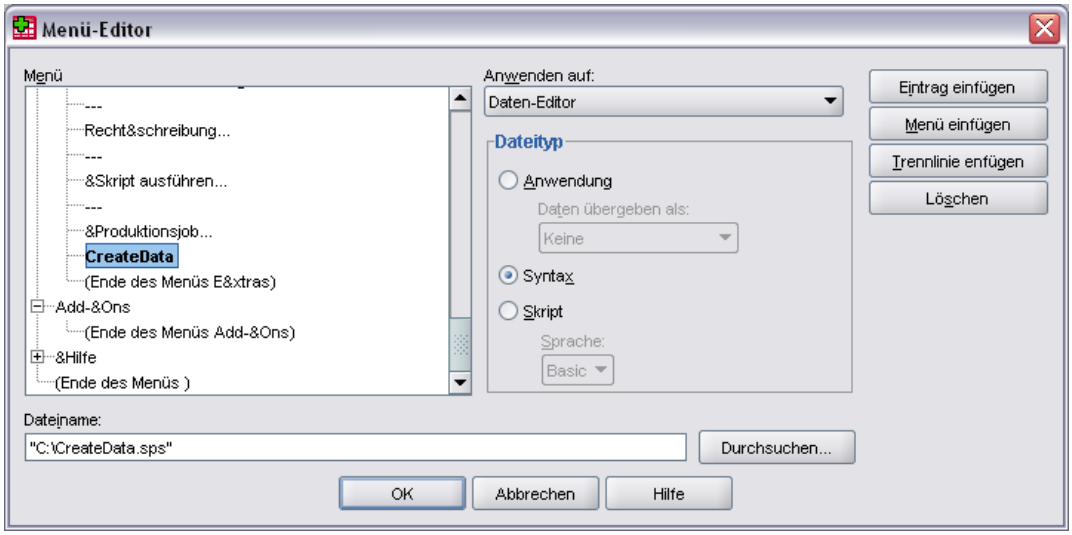
▶ Wählen Sie den Menüeintrag aus, über dem der neue Eintrag eingefügt werden soll.

▶ Klicken Sie zum Einfügen des neuen Menüeintrags auf Eintrag einfügen.

▶ Wählen Sie den Dateityp für den neuen Eintrag aus. Hierbei haben Sie die Auswahl aus Skriptdatei, Befehlssyntax-Datei und externer Anwendung.

▶ Klicken Sie auf Durchsuchen und wählen Sie die Datei aus, die dem Menüeintrag zugewiesen werden soll.

Abbildung 46-1
Dialogfeld "Menü-Editor"



Sie können außerdem vollständig neue Menüs erstellen und Trennlinien zwischen Menüeinträgen einfügen. Wahlweise können Sie festlegen, dass der Inhalt des Daten-Editors automatisch an eine andere Anwendung gesendet wird, wenn Sie diese Anwendung aus dem Menü auswählen.

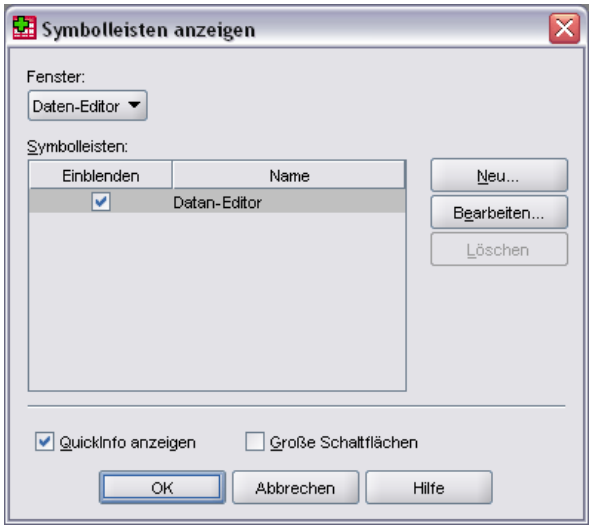
Anpassen von Symbolleisten

Sie können die Symbolleisten von SPSS anpassen und neue Symbolleisten erstellen. Symbolleisten können Symbole für alle in SPSS verfügbaren Funktionen enthalten. Dies schließt Symbole für alle über Menüs verfügbaren Aktionen ein. Außerdem können Symbole enthalten sein, mit denen andere Anwendungen gestartet sowie Befehlssyntax-Dateien und Skriptdateien ausgeführt werden.

Symbolleisten anzeigen

Im Dialogfeld "Symbolleisten anzeigen" können Symbolleisten ein- bzw. ausgeblendet und angepaßt sowie neue Symbolleisten erstellt werden. Symbolleisten können Symbole für alle in SPSS verfügbaren Funktionen enthalten. Dies schließt Symbole für alle über Menüs verfügbaren Aktionen ein. Außerdem können Symbole enthalten sein, mit denen andere Anwendungen gestartet sowie Befehlssyntax-Dateien und Skriptdateien ausgeführt werden.

Abbildung 46-2
Dialogfeld "Symbolleisten anzeigen"



So passen Sie Symbolleisten an:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Symbolleisten
  Anpassen

- Wählen Sie die Symbolleiste aus, die Sie anpassen möchten, und klicken Sie auf Bearbeiten. Zum Erstellen einer neuen Symbolleiste klicken Sie auf Neu.
- Geben Sie bei einer neuen Symbolleiste einen Namen für die Symbolleiste ein, wählen Sie die Fenster aus, in denen die Symbolleiste angezeigt werden soll, und klicken Sie auf Bearbeiten.
- Wenn Sie aus der Liste "Kategorien" eine Kategorie auswählen, werden die in dieser Kategorie verfügbaren Symbole angezeigt.
- Ziehen Sie die gewünschten Symbole auf die im Dialogfeld angezeigte Symbolleiste.
- Zum Entfernen eines Symbols aus einer Symbolleiste ziehen Sie das Symbol aus der im Dialogfeld angezeigten Symbolleiste.

  So erstellen Sie ein Symbol zum Öffnen einer Datei oder zum Ausführen einer Befehlssyntax-Datei bzw. eines Skripts:

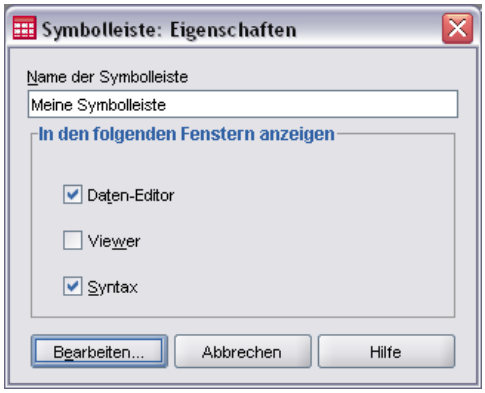
- Klicken Sie im Dialogfeld "Symbolleiste bearbeiten" auf Neues Symbol.
- Geben Sie eine aussagekräftige Beschriftung für das Symbol ein.
- Wählen Sie die gewünschte Aktion für das Symbol aus, also das Öffnen einer Datei oder das Ausführen einer Befehlssyntax-Datei bzw. eines Skripts.
- Klicken Sie auf Durchsuchen und wählen Sie die Datei oder die Anwendung aus, die dem Symbol zugeordnet werden soll.

Neue Symbole werden in der Kategorie “Benutzerdefiniert” angezeigt. Hier finden Sie außerdem die benutzerdefinierten Menüeinträge.

Symbolleiste: Eigenschaften

Verwenden Sie das Dialogfeld “Symbolleiste: Eigenschaften”, um auszuwählen, in welchen Fenstern die ausgewählte Symbolleiste angezeigt werden soll. In diesem Dialogfeld können außerdem Namen für neue Symbolleisten eingegeben werden.

Abbildung 46-3
Dialogfeld “Symbolleiste: Eigenschaften”



So legen Sie die Eigenschaften von Symbolleisten fest:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Ansicht
  Symbolleisten
  Anpassen

- Klicken Sie bei vorhandenen Symbolleisten zuerst auf Bearbeiten und dann im Dialogfeld “Symbolleiste bearbeiten” auf Eigenschaften.

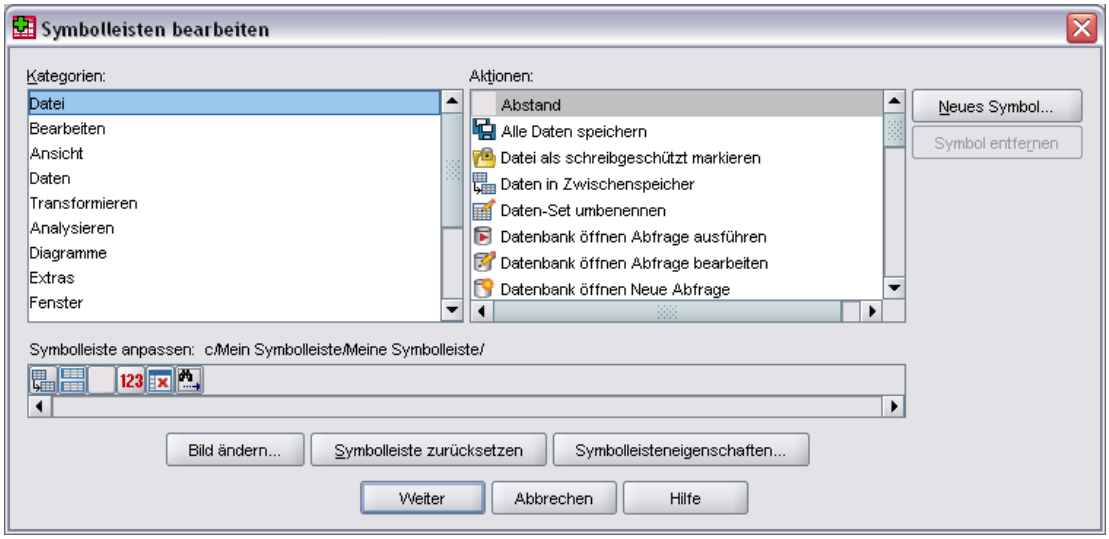
- Klicken Sie bei einer neuen Symbolleiste auf Neues Symbol.

- Wählen Sie die Fenster aus, in denen die Symbolleiste angezeigt werden soll. Geben Sie für eine neue Symbolleiste außerdem einen Namen ein.

Symbolleiste bearbeiten

Verwenden Sie das Dialogfeld “Symbolleiste bearbeiten” zum Anpassen von vorhandenen Symbolleisten und zum Erstellen von neuen Symbolleisten. Symbolleisten können Symbole für alle in SPSS verfügbaren Funktionen enthalten. Dies schließt Symbole für alle über Menüs verfügbaren Aktionen ein. Außerdem können Symbole enthalten sein, mit denen andere Anwendungen gestartet sowie Befehlssyntax-Dateien und Skriptdateien ausgeführt werden.

Abbildung 46-4
Dialogfeld "Symbolleiste bearbeiten"



So ändern Sie Symbole in der Symbolleiste:

- ▶ Wählen Sie das Werkzeug, dessen Symbol Sie in der Symbolleiste ändern möchten.
- ▶ Klicken Sie auf Bild ändern.
- ▶ Wählen Sie die Bilddatei aus, die Sie für das Werkzeug verwenden möchten. Die folgenden Bildformate werden unterstützt: BMP, PNG, GIF, JPG.
  - ■ Die Bilder sollten quadratisch sein. Bilder, die nicht quadratisch sind, werden auf ein quadratisches Format zugeschnitten.
  - ■ Die Größe der Bilder wird automatisch angepasst. Für eine optimale Darstellung sollten Sie Bilder mit 16x16 Pixel für kleine Symbole bzw. Bilder mit 32x32 Pixel für große Symbole in der Symbolleiste verwenden.

Neues Symbol erstellen

Verwenden Sie das Dialogfeld "Neues Symbol erstellen" zum Erstellen von Symbolen, mit denen Sie andere Anwendungen starten und Befehlssyntax-Dateien sowie Skriptdateien ausführen können.

Abbildung 46-5
Dialogfeld "Neues Symbol erstellen"

Kapitel
47

Produktionsjobs

Mithilfe von Produktionsjobs ist eine automatisierte Ausführung von SPSS möglich. Das Programm läuft dann bedienungsfrei ab und wird nach Ausführen des letzten Befehls beendet, sodass Sie sich in der Zwischenzeit anderen Aufgaben widmen oder die automatische Ausführung des Produktionsjobs zu festgesetzten Zeiten planen können. Produktionsjobs bieten sich an, wenn Sie oft dieselben zeitaufwendigen Analysen durchführen müssen, beispielsweise für wöchentliche Berichte.

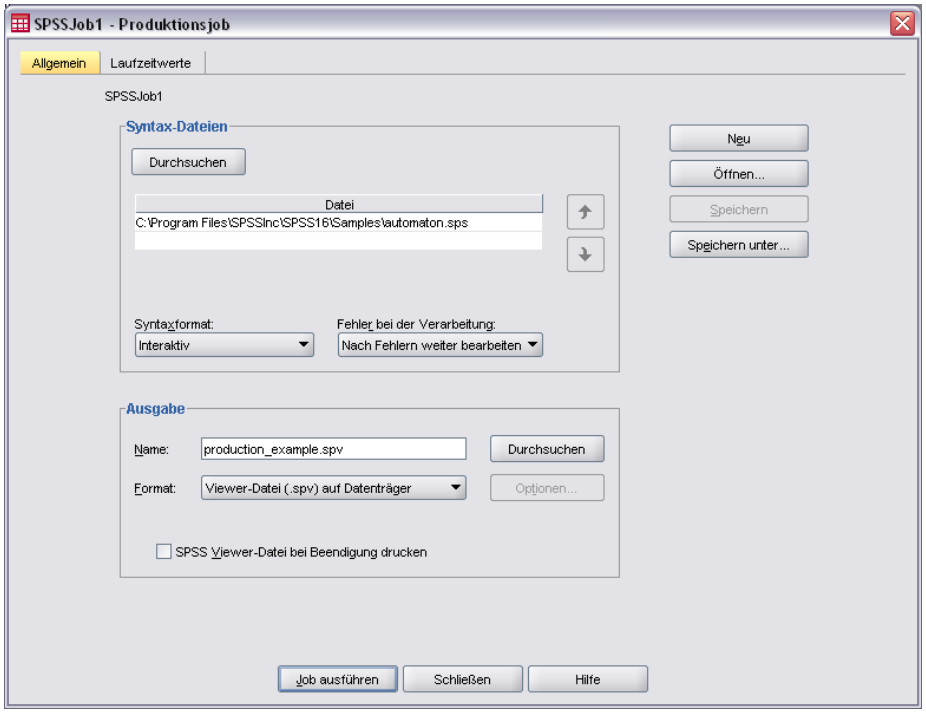
Bei Produktionsjobs werden die Befehle über Befehlssyntax-Dateien an SPSS übermittelt. Bei Befehlssyntax-Dateien handelt es sich um einfache Textdateien, die Befehlssyntax enthalten. Zum Erstellen der Datei können Sie einen beliebigen Text-Editor verwenden. Sie können Befehlssyntax auch erstellen, indem Sie die in einem Dialogfeld getroffene Auswahl in ein Syntax-Fenster einfügen oder die Journaldatei bearbeiten. Für weitere Informationen siehe Arbeiten mit der Befehlssyntax in Kapitel 12 auf S. 269.

So erstellen Sie einen Produktionsjob:

- Wählen Sie in einem beliebigen Fenster die folgenden Optionen aus den Menüs aus:

  Extras
  Produktionsjob

Abbildung 47-1
Dialogfeld "Produktionsjob"



▶ Klicken Sie auf Neu, um einen neuen Produktionsjob zu erstellen, bzw. auf Öffnen, um einen bestehenden Produktionsjob zu öffnen.

Anmerkung: Produktionsmodusjob-Dateien (.spp), die in Versionen vor 16.0 erstellt wurden, funktionieren in Version 16.0 und höher nicht. Es ist ein Konvertierungsdienstprogramm verfügbar, mit dem Produktionsmodusjob-Dateien unter Windows und Macintosh in Produktionsjobs (.spj) umgewandelt werden können. Für weitere Informationen siehe Konvertieren von Produktionsmodus-Dateien auf S. 522.

▶ Wählen Sie eine oder mehrere Befehlssyntax-Dateien aus.

Anmerkung: Im Modus für verteilte Analysen müssen die Speicherstellen der Syntaxdateien manuell eingegeben werden und relative Pfade sind relativ zum Remote-Server. Für weitere Informationen siehe Absolute und relative Pfadangaben in Kapitel 4 auf S. 78.

▶ Wählen Sie Namen, Speicherort und Format der Ausgabedatei aus.

▶ Klicken Sie auf Speichern bzw. Speichern unter, um den Produktionsjob zu speichern.

Syntaxformat. Steuert die Form der für den Job verwendeten Syntaxregeln.

- **Interaktiv.** Jeder Befehl muss mit einem Punkt enden. Punkte können an einer beliebigen Stelle im Befehl auftreten, und Befehle können sich über mehrere Zeilen erstrecken. Ein Punkt als letztes Zeichen in einer Zeile wird jedoch stets als Ende des Befehls interpretiert. Fortgesetzte Zeilen und neue Befehle können an einer beliebigen Position in einer neuen

Zeile beginnen. Diese "interaktiven" Regeln sind in Kraft, wenn Sie Befehle in einem Syntaxfenster auswählen und ausführen.

- **Stapel.** Jeder Befehl muss am Anfang einer neuen Zeile beginnen (Leerzeichen vor dem Befehl sind nicht zulässig); fortgesetzte Zeilen müssen um mindestens ein Leerzeichen eingerückt sein. Sollen neue Befehle eingerückt werden, geben Sie ein Pluszeichen, einen Bindestrich oder einen Punkt als erstes Zeichen am Anfang der Zeile ein, und rücken Sie dann den eigentlichen Befehl nach Wunsch ein. Der Punkt am Ende des Befehls ist optional. Diese Einstellung ist kompatibel mit den Syntaxregeln für Befehlsdateien im Befehl `INCLUDE`.

  Anmerkung: Verwenden Sie die Option "Batch" *nicht*, wenn Ihre Syntaxdateien `GGRAPH`-Befehlssyntax mit `GPL`-Anweisungen enthalten. `GPL`-Anweisungen werden nur bei interaktiven Regeln ausgeführt.

Fehler bei der Verarbeitung. Steuert die Behandlung von Fehlerbedingungen im Job.

- **Nach Fehlern weiter bearbeiten.** Fehler im Job führen nicht automatisch dazu, dass die Befehlsverarbeitung abgebrochen wird. Die Befehle in den Produktionsjob-Dateien werden als Teil des normalen Befehlsstroms behandelt, und die Befehlsverarbeitung wird normal fortgesetzt.
- **Verarbeitung sofort anhalten.** Die Befehlsverarbeitung wird angehalten, sobald der erste Fehler in einer Produktionsjob-Datei auftritt. Diese Einstellung ist kompatibel mit dem Verhalten von Befehlsdateien im Befehl `INCLUDE`.

Ausgabe. Dient zur Festlegung von Name, Speicherort und Format der Ergebnisse von Produktionsjobs. Die folgenden Formatoptionen sind verfügbar:

- **Viewer-Datei (.spv) auf Datenträger.** Die Ergebnisse werden im SPSS Viewer-Format am angegebenen Dateispeicherort gespeichert.
- **Viewer-Datei (.spv) nach PES Repository.** Dafür ist die Option "SPSS Adaptor für Enterprise Services" erforderlich.
- **Webberichte (.spw) nach PES Repository.** Dafür ist die Option "SPSS Adaptor für Enterprise Services" erforderlich.
- **Word/RTF (*.doc).** Pivot-Tabellen werden mit sämtlichen Formatierungsattributen wie Zellenrahmen, Schriftarten und Hintergrundfarben, als Word-Tabellen exportiert. Textausgaben werden als formatierter RTF-Text exportiert. Unter Windows-Betriebssytemen werden Diagramme im Format EMF (Enhanced Metafile) in das Dokument aufgenommen. Unter anderen Betriebssystemen werden Diagramme im Format PNG aufgenommen. Textausgaben werden immer mit einem nicht proportionalen Zeichensatz (mit festem Abstand) angezeigt und mit denselben Schriftartenattributen exportiert. Für die richtige Ausrichtung von durch Leerzeichen getrennten Textausgaben ist ein nicht proportionaler Zeichensatz (mit festem Abstand) erforderlich.

  Anmerkung: Besonders breite Tabellen werden von Microsoft Word möglicherweise nicht ordnungsgemäß angezeigt.
- **Excel (*.xls).** Die Zeilen, Spalten und Zellen von Pivot-Tabellen werden mit sämtlichen Formatierungsattributen wie Zellenrahmen, Schriftarten, Hintergrundfarben als Excel-Zeilen, -Spalten und -Zellen exportiert. Textausgaben werden mit allen Schriftartattributen exportiert. Jede Zeile in der Textausgabe entspricht einer Zeile in der Excel-Datei, wobei der gesamte Inhalt der Zeile in einer einzelnen Zelle enthalten ist. Diagramme werden nicht mit aufgenommen.

- **HTML (*.htm).** Pivot-Tabellen werden als HTML-Tabellen exportiert. Textausgaben werden als vorformatierter HTML-Text exportiert. Diagramme werden als Verweis eingebettet. Daher sollten Sie Diagramme in einem für die Aufnahme in HTML-Dokumente geeigneten Format exportieren (z. B. PNG oder JPEG).
- **Portable Document Format (*.pdf).** Alle Ausgaben werden so exportiert, wie sie in der Druckvorschau/Seitenansicht angezeigt werden. Alle Formatierungsattribute bleiben erhalten.
- **PowerPoint file (*.ppt).** Pivot-Tabellen werden als Word-Dateien exportiert und sind auf separaten Folien in der PowerPoint-Datei eingebettet (je eine Pivot-Tabelle auf einer Folie). Sämtliche Formatierungsattribute der Pivot-Tabelle (z. B. Zellenrahmen, Schriftarten und Hintergrundfarben) werden beibehalten. Diagramme werden im Format TIFF exportiert. Textausgaben sind nicht eingeschlossen.

  Anmerkung: Der Export nach PowerPoint ist nur unter Windows-Betriebssystemen und nicht in der Studentenversion verfügbar.
- **Text (*.txt).** Zu den Textausgabeformaten gehören einfacher Text, UTF-8 und UTF-16. Pivot-Tabellen können als durch Tabulatoren getrennter Text oder als durch Leerzeichen getrennter Text exportiert werden. Alle Textausgaben werden in durch Leerzeichen getrenntem Format exportiert. Bei Diagrammen wird in der Textdatei für jedes Diagramm eine Zeile mit der Angabe des Dateinamens für das exportierte Diagramm eingefügt.

SPSS Viewer-Datei bei Beendigung drucken. Sendet nach Abschluss des Produktionsjobs die endgültige Viewer-Ausgabedatei an den Drucker.

Job ausführen. Damit wird der Produktionsjob in einer separaten Sitzung ausgeführt. Dies bietet sich insbesondere an, um neue Produktionsjobs vor der Bereitstellung zu testen.

Produktionsjobs mit OUTPUT-Befehlen

Produktionsjobs berücksichtigen SPSS `OUTPUT`-Befehle, wie beispielsweise `OUTPUT SAVE`, `OUTPUT ACTIVATE` und `OUTPUT NEW`. Die im Verlauf eines Produktionsjobs ausgeführten `OUTPUT SAVE`-Befehle schreiben die Inhalte der angegebenen Ausgabedokumente in die angegebenen Speicherstellen. Dies erfolgt zusätzlich zu der durch den Produktionsjob erstellten Ausgabedatei. Wenn Sie mithilfe von `OUTPUT NEW` ein neues Ausgabedokument erstellen, sollten Sie es explizit mit dem Befehl `OUTPUT SAVE` speichern.

Die Ausgabedatei eines Produktionsjobs besteht aus dem Inhalt des aktiven Ausgabedokuments zum Zeitpunkt der Beendigung des Jobs. Bei Jobs mit `OUTPUT`-Befehlen enthält die Ausgabedatei möglicherweise nicht die gesamte Ausgabe, die in der Sitzung erstellt wurde. Beispiel: Angenommen, der Produktionsjob besteht aus einer Reihe von SPSS-Prozeduren, gefolgt von einem `OUTPUT NEW`-Befehl, gefolgt von weiteren SPSS-Prozeduren, jedoch ohne weitere `OUTPUT`-Befehle. Der `OUTPUT NEW`-Befehl definiert ein neues aktives Ausgabedokument. Am Ende des Produktionsjobs enthält es nur Ausgaben aus den Prozeduren, die nach dem `OUTPUT NEW`-Befehl ausgeführt wurden.

HTML-Optionen

Tabellenoptionen. Es sind keine Tabellenoptionen für das HTML-Format verfügbar. Alle Pivot-Tabellen werden in HTML-Tabellen konvertiert.

Bildoptionen. Die folgenden Bildtypen stehen zur Verfügung: EPS, JPEG, TIFF, PNG und BMP. Unter Windows-Betriebssystemen ist außerdem das Format EMF (Enhanced Metafile, erweiterte Metadatei) verfügbar. Außerdem können Sie das Bild von 1 % bis 200 % skalieren.

PowerPoint-Optionen

Tabellenoptionen. Sie können die Einträge in der Viewer-Gliederung als Folientitel verwenden. Jede Folie enthält ein einzelnes Ausgabeobjekt. Der Titel wird aus dem Gliederungseintrag für das Element im Gliederungsfenster des Viewers gebildet.

Bildoptionen. Sie können das Bild von 1 % bis 200 % skalieren. (Alle Bilder werden im TIFF-Format nach PowerPoint exportiert.)

Anmerkung: Das PowerPoint-Format steht nur auf Windows-Betriebssystemen zur Verfügung. Außerdem ist PowerPoint 97 oder höher erforderlich.

PDF-Optionen

Lesezeichen einbetten. Mit dieser Option werden Lesezeichen in das PDF-Dokument aufgenommen, die den Einträgen in der Viewer-Gliederung entsprechen. Wie das Viewer-Gliederungsfenster können auch Lesezeichen die Navigation in Dokumenten mit einer Vielzahl an Ausgabeobjekten erheblich erleichtern.

Schriftarten einbetten. Durch das Einbetten von Schriftarten wird sichergestellt, dass das PDF-Dokument auf allen Computern gleich dargestellt wird. Anderenfalls kann es, wenn im Dokument verwendete Schriftarten auf dem Computer, der zur Anzeige (oder zum Drucken) des PDF-Dokuments verwendet wird, nicht zur Verfügung stehen, durch Schriftartenersetzung zu suboptimalen Ergebnissen kommen.

Text-Optionen

Tabellenoptionen. Pivot-Tabellen können als durch Tabulatoren getrennter Text oder als durch Leerzeichen getrennter Text exportiert werden. Beim leerzeichengetrennten Format können Sie außerdem folgende Eigenschaften festlegen:

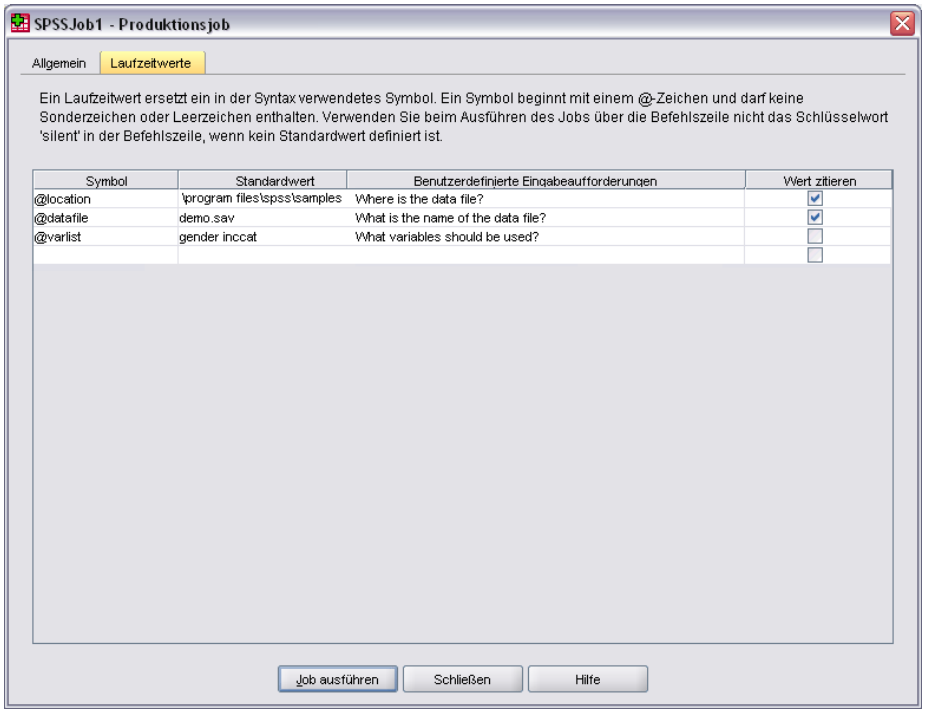
- **Spaltenbreite.** Mit Automatisch anpassen werden keine Spalteninhalte umgebrochen und jede Spalte ist so breit, wie das längste Label bzw. der längste Wert in der Spalte. Mit Benutzerdefiniert wird eine maximale Spaltenbreite festgelegt, die für alle Spalten in der Tabelle gilt. Bei Werten, die breiter sind, wird ein Zeilenumbruch durchgeführt, sodass sie sich auch auf die nächste Zeile in der betreffenden Spalte erstrecken.
- **Zeilen-/Spaltenbegrenzungszeichen.** Legt die Zeichen fest, die für Zeilen- und Spaltenrahmen verwendet werden. Um die Anzeige von Zeilen- und Spaltenrahmen zu unterdrücken, geben Sie als Werte Leerzeichen ein.

Bildoptionen. Die folgenden Bildtypen stehen zur Verfügung: EPS, JPEG, TIFF, PNG und BMP. Unter Windows-Betriebssystemen ist außerdem das Format EMF (Enhanced Metafile, erweiterte Metadatei) verfügbar. Außerdem können Sie das Bild von 1 % bis 200 % skalieren.

Laufzeitwerte

Die in Produktionsjob-Dateien definierten und in Befehlssyntax-Dateien verwendeten Laufzeitwerte vereinfachen Aufgaben, wie zum Beispiel das Durchführen derselben Analyse für verschiedene Datendateien oder das Ausführen derselben Befehle für verschiedene Variablen-Sets. So können Sie beispielsweise den Laufzeitwert *@datdtei* definieren, sodass bei jedem Ausführen von Produktionsjobs, die anstelle eines Namens für die Datendatei in der Befehlssyntax-Datei den String *@datdtei* enthalten, ein Name für die Datendatei abgefragt wird.

Abbildung 47-2
Dialogfeld "Produktionsjob", Registerkarte "Laufzeitwerte"



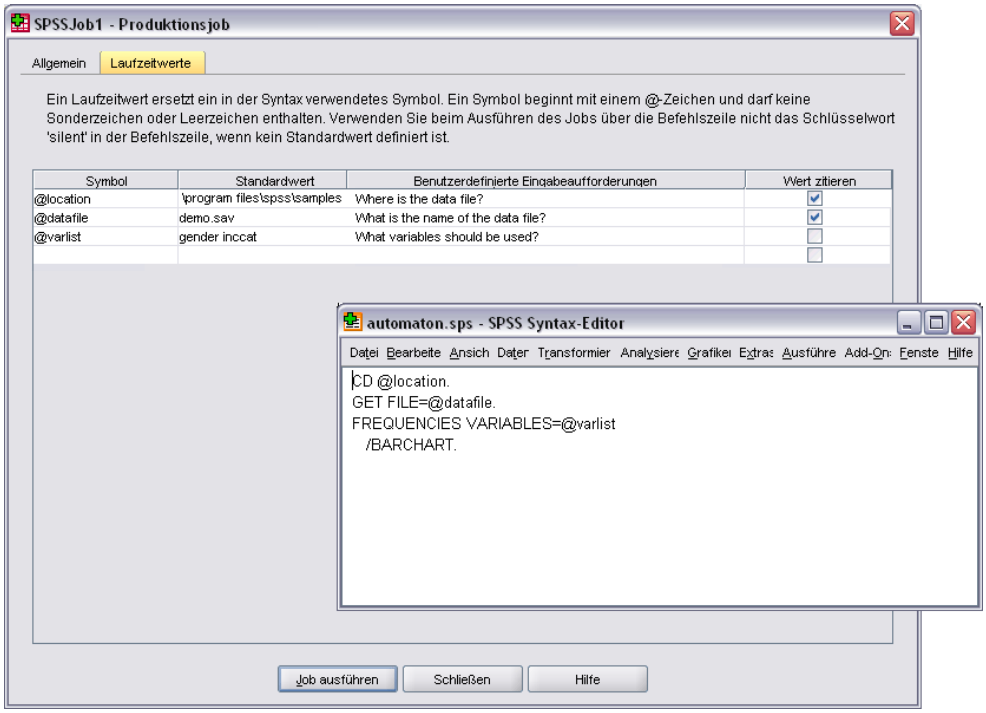
Symbol. Der String in der Befehlssyntax-Datei, der den Produktionsjob auslöst, um den Benutzer zur Eingabe eines Werts aufzufordern. Der Name des Symbols muss mit einem @-Zeichen beginnen und den Regeln zum Benennen von Variablen entsprechen. Für weitere Informationen siehe Variablennamen in Kapitel 5 auf S. 83.

Standardwert. Dies ist der Wert, der vom Produktionsjob in der Standardeinstellung vorgegeben wird, falls Sie keinen anderen Wert eingeben. Dieser Wert wird angezeigt, wenn Sie beim Ausführen des Produktionsjobs zum Eingeben von Informationen aufgefordert werden. Sie können den Wert während der Ausführung des Programms ersetzen oder modifizieren. Wenn Sie keinen Standardwert angeben, dürfen Sie das Schlüsselwort `silent` nicht verwenden, wenn Sie den Produktionsjob mit Befehlszeilenschaltern ausführen, es sei denn, sie verwenden auch den Schalter `-symbol` zur Angabe von Laufzeitwerten. Für weitere Informationen siehe Ausführen von Produktionsjobs aus der Befehlszeile auf S. 521.

Benutzerdefinierte Eingabeaufforderungen. Der beschreibende Text, der angezeigt werden soll, wenn der Produktionsjob zur Eingabe von Informationen auffordert. Sie können zum Beispiel mit der Formulierung “Welche Datendatei möchten Sie verwenden?” ein Feld bezeichnen, in das der Name einer Datendatei eingegeben werden muss.

Wert zitieren. Schließt den Standardwert bzw. den vom Benutzer eingegebenen Wert in Anführungsstriche ein. Beispielsweise müssen Werte, die Leerzeichen enthalten, in Anführungszeichen gesetzt werden.

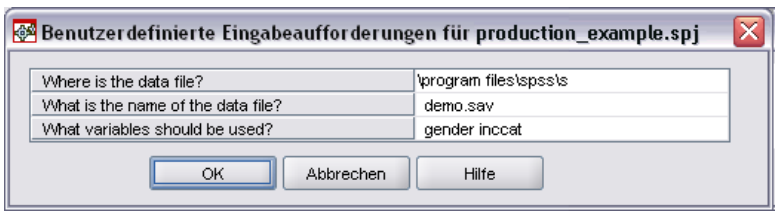
Abbildung 47-3
Laufzeitsymbole in einer Befehlssyntax-Datei



Benutzerdefinierte Eingabeaufforderungen

Sie werden zur Eingabe von Werten aufgefordert, wenn Sie einen Produktionsjob mit definierten Laufzeitsymbolen ausführen. Sie können die angezeigten Standardwerte ersetzen oder ändern. Diese Werte ersetzen dann die Laufzeitsymbole in allen zu dem Produktionsjob gehörenden Befehlssyntax-Dateien.

Abbildung 47-4
Produktion – Dialogfeld “Benutzerdefinierte Eingabeaufforderungen”



Ausführen von Produktionsjobs aus der Befehlszeile

Mit Befehlszeilenschaltern können Sie Produktionsjobs unter Verwendung der unter Ihrem Betriebssystem verfügbaren Dienstprogrammen automatisch zeitgesteuert ausführen. Die Grundform des Befehlszeilenarguments lautet:

```
spss filename.spj -production
```

Je nachdem auf welche Weise Sie den Produktionsjob aufrufen, müssen Sie möglicherweise Verzeichnispfade für die ausführbare *spss*-Datei (diese befindet sich in dem Verzeichnis, in dem die Anwendung installiert wurde) und/oder die Produktionsjob-Datei angeben.

Sie können Produktionsjobs aus einer Befehlszeile heraus mit den folgenden Schaltern ausführen:

-production [prompt|silent]. Startet die Anwendung im Produktionsmodus. Die Schlüsselwörter `prompt` und `silent` geben an, ob das Dialogfeld angezeigt werden soll, das zur Eingabe von Laufzeitwerten angibt, wenn diese im Job angegeben sind. Das Schlüsselwort `prompt` ist die Standardvorgabe. Mit diesem Schlüsselwort wird das Dialogfeld angezeigt. Mit dem Schlüsselwort `silent` wird das Dialogfeld unterdrückt. Wenn Sie das Schlüsselwort `silent` verwenden, können Sie die Laufzeitsymbole mit dem Schalter `-symbol` angeben. Ansonsten wird der Standardwert verwendet. Die Schalter `-switchserver` und `-singleseat` werden bei Verwendung des Schalters `-production` ignoriert.

-symbol <Werte>. Liste der Symbol-/Wert-Paare, die im Produktionsjob verwendet werden. Jeder Symbolname beginnt mit @. Werte, die Leerzeichen enthalten, müssen in Anführungszeichen gesetzt werden. Regeln für die Aufnahme von Anführungszeichen oder Apostrophen in String-Literalen variieren je nach Betriebssystem, aber das Einschließen eines Strings, der einzelne Anführungszeichen oder Apostrophe enthält, in doppelte Anführungszeichen funktioniert normalerweise (Beispiel: `"'ein Wert in Anführungszeichen'"`).

Zur Ausführung von Produktionsjobs auf einem Remote-Server im Modus für verteilte Analysen müssen Sie außerdem die Anmeldeinformationen für den Server angeben:

-server <Inet:Hostname:Port>. Der Name bzw. die IP-Adresse und die Portnummer des Servers. Nur unter Windows.

-user <Name>. Ein gültiger Benutzername. Wenn ein Domänenname erforderlich ist, müssen Sie dem Benutzernamen den Domänennamen und einen umgekehrten Schrägstrich (\) voranstellen. Nur unter Windows.

-password <Passwort>. Das Passwort des Bentutzers. Nur unter Windows.

Beispiel

```
spss \production_jobs\prodjob1.spj -production silent -symbol @datdtei /Daten/Juli-Daten.sav
```

- Bei diesem Beispiel wird davon ausgegangen, dass Sie die Befehlszeile vom Installationsverzeichnis aus ausführen, sodass kein Pfad für die ausführbare *spss*-Datei angegeben werden muss.
- Bei diesem Beispiel wird außerdem davon ausgegangen, dass der Produktionsjob angibt, dass der Wert für *@datdtei* in Anführungszeichen gesetzt werden sollte (Kontrollkästchen “Wert zitieren” auf der Registerkarte “Laufzeitwerte”), sodass bei der Angabe der Datendatei

in der Befehlszeile keine Anführungszeichen erforderlich sind. Anderenfalls müssten Sie beispielsweise '/Daten/Juli-Daten.sav angeben, um Anführungszeichen bei der Angabe der Datendatei zu verwenden, da Dateiangaben in der Befehlssyntax in Anführungszeichen gesetzt werden müssen.

- Für den Verzeichnispfad zum Speicherort des Produktionsjobs wird die Windows-Konvention mit umgekehrten Schrägstrichen verwendet. Unter Microsoft und Linux müssen Sie stattdessen normale Schrägstriche verwenden. Normale Schrägstriche und die Angabe der Datendateien in Anführungsstrichen funktioniert unter allen Betriebssystemen, da dieser in Anführungszeichen gesetzte String in die Befehlszeilensyntax eingefügt wird und unter allen Betriebssystemen normale Schrägstriche in Befehlen zulässig sind, die Dateiangaben enthalten (z. B. `GET FILE`, `GET DATA`, `SAVE`)
- Durch das Schlüsselwort `silent` werden alle Eingabeaufforderungen im Produktionsjob unterdrückt und mit dem Schalter `-symbol` werden Datendateiname und Speicherort in Anführungsstrichen immer dann eingefügt, wenn das Laufzeitsymbol *@datdtei* in den Befehlssyntax-Dateien des Produktionsjobs vorkommt.

Konvertieren von Produktionsmodus-Dateien

Produktionsmodusjob-Dateien (.spp), die in Versionen vor 16.0 erstellt wurden, funktionieren in Version 16.0 und höher nicht. In früheren Versionen unter Windows bzw. Macintosh erstellte Jobdateien des Produktionsmodus können Sie mithilfe von *prodconvert* (im Installationsverzeichnis zu finden) in neue Produktionsjob-Dateien (.spj) umwandeln. Die Grundspezifikation (zugleich die einzige Spezifikation) lautet:

```
prodconvert dateiname.spp
```

Eine neue Datei mit demselben Namen, jedoch mit der Erweiterung *.spj* wird im selben Verzeichnis wie die Originaldatei erstellt.

Einschränkungen

- Die Formate von WMF- und EMF-Diagrammen werden nicht unterstützt. Anstelle dieser Formate wird das Format PNG verwendet.
- Die Exportoptionen Ausgabedokument (ohne Diagramme), Nur Diagramme und Nichts werden nicht unterstützt. Alle Ausgabeobjekte, die vom ausgewählten Format unterstützt werden, werden berücksichtigt.
- Die Option, mit der SPPS angezeigt und Abschluss des Jobs geöffnet gelassen werden kann, wird nicht unterstützt Die Produktionssitzung wird stets nach Abschluss des Jobs geschlossen.
- Die Einstellungen für den Remote-Server werden ignoriert. Um Remote-Server-Einstellungen für die verteilte Analyse anzugeben, müssen Sie den Produktionsjob über eine Befehlszeile ausführen und dabei die Servereinstellungen mithilfe von Befehlszeilenschaltern angeben. Für weitere Informationen siehe Ausführen von Produktionsjobs aus der Befehlszeile auf S. 521.
- Einstellungen zur Veröffentlichung im Web werden ignoriert.

Kapitel
48

Ausgabeverwaltungssystem (OMS)

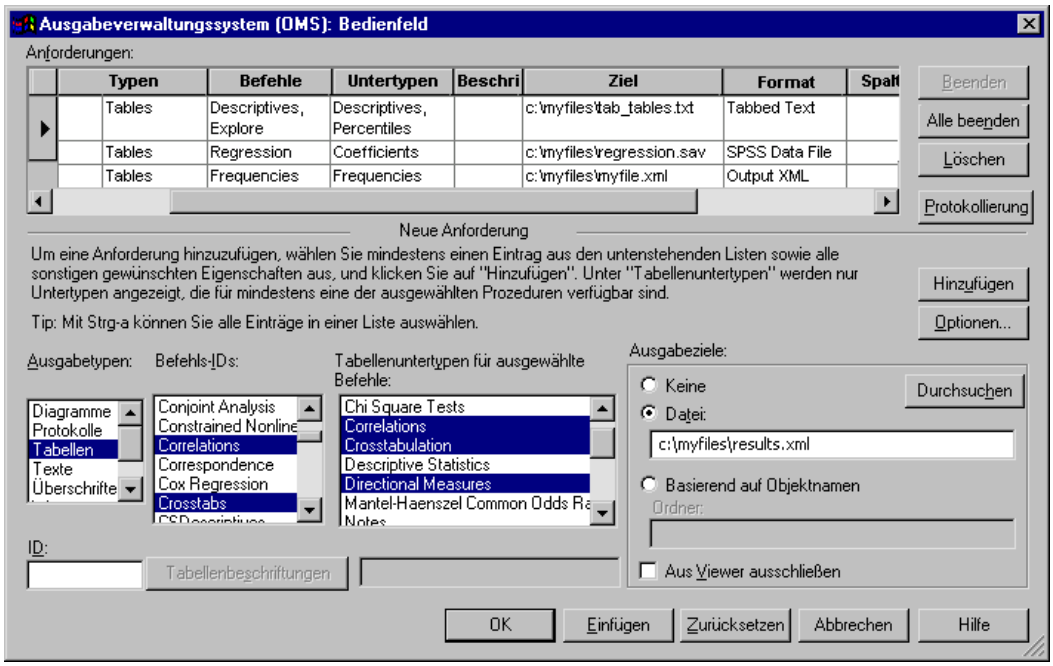
Das Ausgabeverwaltungssystem (Output Management System, OMS) bietet die Möglichkeit, ausgewählte Ausgabekategorien automatisch in verschiedene Ausgabedateien in unterschiedlichen Formaten schreiben zu lassen. Zu diesen Formaten gehören: SPSS-Datendateiformat (.sav), SPSS-Viewer-Dateiformat (.spv), XML, HTML und Text. Für weitere Informationen siehe OMS: Optionen auf S. 529.

So verwenden Sie das Bedienfeld des Ausgabeverwaltungssystems:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Extras
  OMS-Steuerung...

Abbildung 48-1
Systemsteuerung des Ausgabeverwaltungssystems



Mit dem Bedienfeld können Sie die Weiterleitung der Ausgaben an verschiedene Ziele starten und beenden.

- Alle OMS-Anforderungen bleiben aktiv, bis sie ausdrücklich beendet werden oder die Sitzung endet.

- Eine in einer OMS-Anforderung angegebene Zieldatei steht für andere Prozeduren und andere Anwendungen so lange nicht zur Verfügung, bis die OMS-Anforderung beendet ist.
- Wenn eine OMS-Anforderung aktiv ist, werden die angegebenen Zieldateien im Arbeitsspeicher (RAM) abgelegt. Aktive OMS-Anforderungen, bei denen große Ausgabemengen in externe Dateien geschrieben werden, können somit große Mengen an Arbeitsspeicher belegen.
- Mehrere gleichzeitig ausgeführte OMS-Anforderungen sind unabhängig voneinander. Eine Ausgabe kann in unterschiedlichen Formaten an verschiedene Positionen weitergeleitet werden, je nach den Spezifikationen in den einzelnen OMS-Anforderungen.
- Die Ausgabeobjekte werden in der Reihenfolge am jeweiligen Ziel ausgegeben, in der sie erstellt wurden. Diese Reihenfolge bei der Erstellung ergibt sich aus der Reihenfolge und der Nutzung der Prozedur, die die Ausgabe erzeugt.

So fügen Sie neue OMS-Anforderungen hinzu:

▶ Wählen Sie die zu berücksichtigenden Ausgabetypen aus (z. B. Tabellen, Diagramme usw.). Für weitere Informationen siehe Ausgabeobjekttypen auf S. 526.

▶ Wählen Sie die einzuschließenden Befehle aus. Soll die gesamte Ausgabe berücksichtigt werden, wählen Sie alle Elemente in der Liste aus. Für weitere Informationen siehe Befehls-IDs und Tabellenuntertypen auf S. 527.

▶ Bei Befehlen, die eine Pivot-Tabelle als Ausgabe erzeugen, wählen Sie die zugehörigen Tabellentypen aus.

Die Liste enhält nur die Tabellen, die in den ausgewählten Befehlen zur Verfügung stehen. Es werden sämtliche Tabellentypen aufgeführt, die in mindestens einem der ausgewählten Befehle in der Liste verfügbar sind. Falls Sie keine Befehle ausgewählt haben, werden alle Tabellentypen aufgeführt. Für weitere Informationen siehe Befehls-IDs und Tabellenuntertypen auf S. 527.

▶ Wenn Sie Tabellen nicht nach den Untertypen, sondern nach dem Beschriftungstext auswählen möchten, klicken Sie auf Beschriftungen. Für weitere Informationen siehe Labels auf S. 528.

▶ Klicken Sie auf Optionen, und bestimmen Sie das Ausgabeformat (z. B. SPSS-Datendatei, XML, HTML). Standardmäßig wird das Output XML-Format verwendet. Für weitere Informationen siehe OMS: Optionen auf S. 529.

▶ Geben Sie ein Ausgabeziel an:

- **Datei.** Alle ausgewählten Ausgaben werden in eine Datei weitergeleitet.
- **Basierend auf Objektnamen.** Die Ausgabe wird anhand der Objektnamen an mehrere Zieldateien weitergeleitet. Für jedes Ausgabeobjekt wird eine separate Datei erstellt. Die Dateinamen basieren entweder auf den Namen von Tabellenuntertypen oder den Beschriftungen der Tabellen. Geben Sie den Namen des Zielordners ein.
- **Neues Daten-Set.** Bei Ausgaben im Format von SPSS-Datendateien können Sie die Ausgabe in ein Daten-Set weiterleiten. Das Daten-Set ist für die anschließende Verwendung in derselben Sitzung verfügbar. Es wird jedoch nicht gespeichert, sofern Sie es nicht ausdrücklich vor dem Beenden der Sitzung als Datei speichern. Diese Option ist nur für Ausgaben im Format von SPSS-Datendateien verfügbar. Die Namen von Daten-Sets müssen den Regeln zum

Benennen von Variablen entsprechen. Für weitere Informationen siehe Variablennamen in Kapitel 5 auf S. 83.

- Die folgenden Optionen sind verfügbar:
  - Schließen Sie die ausgewählten Ausgaben aus dem Viewer aus. Mit Aus Viewer ausschließen werden die Ausgabetypen in der OMS-Anforderung nicht im Viewer-Fenster angezeigt. Enthalten mehrere aktive OMS-Anforderungen dieselben Ausgabetypen, wird anhand der jüngsten OMS-Anforderung mit den betreffenden Ausgabetypen entschieden, ob die Ausgabetypen angezeigt werden sollen oder nicht. Für weitere Informationen siehe Ausschließen der Ausgabeanzeige aus dem Viewer auf S. 534.
  - Weisen Sie der Anfrage einen ID-String zu. Alle Anfragen erhalten automatisch einen ID-Wert, und Sie können den Standard-ID-String durch eine aussagekräftige ID ersetzen. Dies ist insbesondere dann von Nutzen, wenn mehrere aktive Anfragen vorliegen, die Sie einfach unterscheiden möchten. Die zugewiesenen ID-Werte dürfen nicht mit einem Dollar-Zeichen ($) beginnen.

Hier finden Sie einige Tipps, wie Sie mehrere Elemente in einer Liste auswählen:

- Drücken Sie Strg+A, um alle Elemente in einer Liste gleichzeitig auszuwählen.
- Wenn Sie bei gedrückter Umschalttaste klicken, können Sie mehrere aufeinander folgende Elemente auswählen.
- Wenn Sie bei gedrückter Strg-Taste klicken, können Sie mehrere nicht aufeinander folgende Elemente auswählen.

So beenden und löschen Sie OMS-Anforderungen:

Aktive und neue OMS-Anforderungen werden in der Liste “Anforderungen” aufgeführt. Die neueste Anfrage befindet sich dabei an oberster Stelle. Die Breite der Datenspalten kann geändert werden. Klicken Sie hierzu auf die Spaltenbegrenzungen, und ziehen Sie sie an die gewünschte Position. Darüber hinaus können Sie in horizontaler Richtung durch die Liste blättern und so weitere Informationen zu einer bestimmten Anfrage anzeigen lassen.

Ein Sternchen (*) neben dem Wort Aktiv in der Spalte *Status* bedeutet, dass die zugehörige OMS-Anforderung mit Befehlssyntax erzeugt wurde, die im Bedienfeld nicht zur Verfügung steht.

So beenden Sie eine bestimmte aktive OMS-Anforderung:

- Klicken Sie in der Liste “Anforderungen” auf eine beliebige Zelle in der Zeile dieser Anfrage.
- Klicken Sie auf Beenden.

So beenden Sie alle aktiven OMS-Anforderungen:

- Klicken Sie auf Alle beenden.

So löschen Sie eine neue Anfrage (eine hinzugefügte Anfrage, die noch nicht aktiv ist):

- Klicken Sie in der Liste “Anforderungen” auf eine beliebige Zelle in der Zeile dieser Anfrage.
- Klicken Sie auf Löschen.

Anmerkung: Aktive OMS-Anforderungen werden erst dann beendet, wenn Sie auf OK klicken.

Ausgabeobjekttypen

Es gibt sieben verschiedene Arten von Ausgabeobjekten:

Diagramme. Diagrammobjekte werden nur bei den Zielformaten XML und HTML berücksichtigt. Beim HTML-Format werden die Bilddateien in einem separaten Unterverzeichnis (Ordner) gespeichert.

Protokolle. Protokolltextobjekte. Protokollobjekte enthalten bestimmte Arten von Fehler- und Warnmeldungen. Je nach den Einstellungen unter "Optionen" (Menü "Bearbeiten", "Optionen", Registerkarte "Viewer") umfassen die Log-Objekte möglicherweise auch die Befehlssyntax, die während der Sitzung ausgeführt wurde. Protokollobjekte werden im Gliederungsfenster des Viewers mit *Log* gekennzeichnet.

Tabellen. Ausgabeobjekte, die im Viewer als Pivot-Tabellen dargestellt werden (einschließlich der Anmerkungstabellen). Tabellen sind die einzigen Ausgabeobjekte, die im SPSS-Datendateiformat (.sav) weitergeleitet werden können.

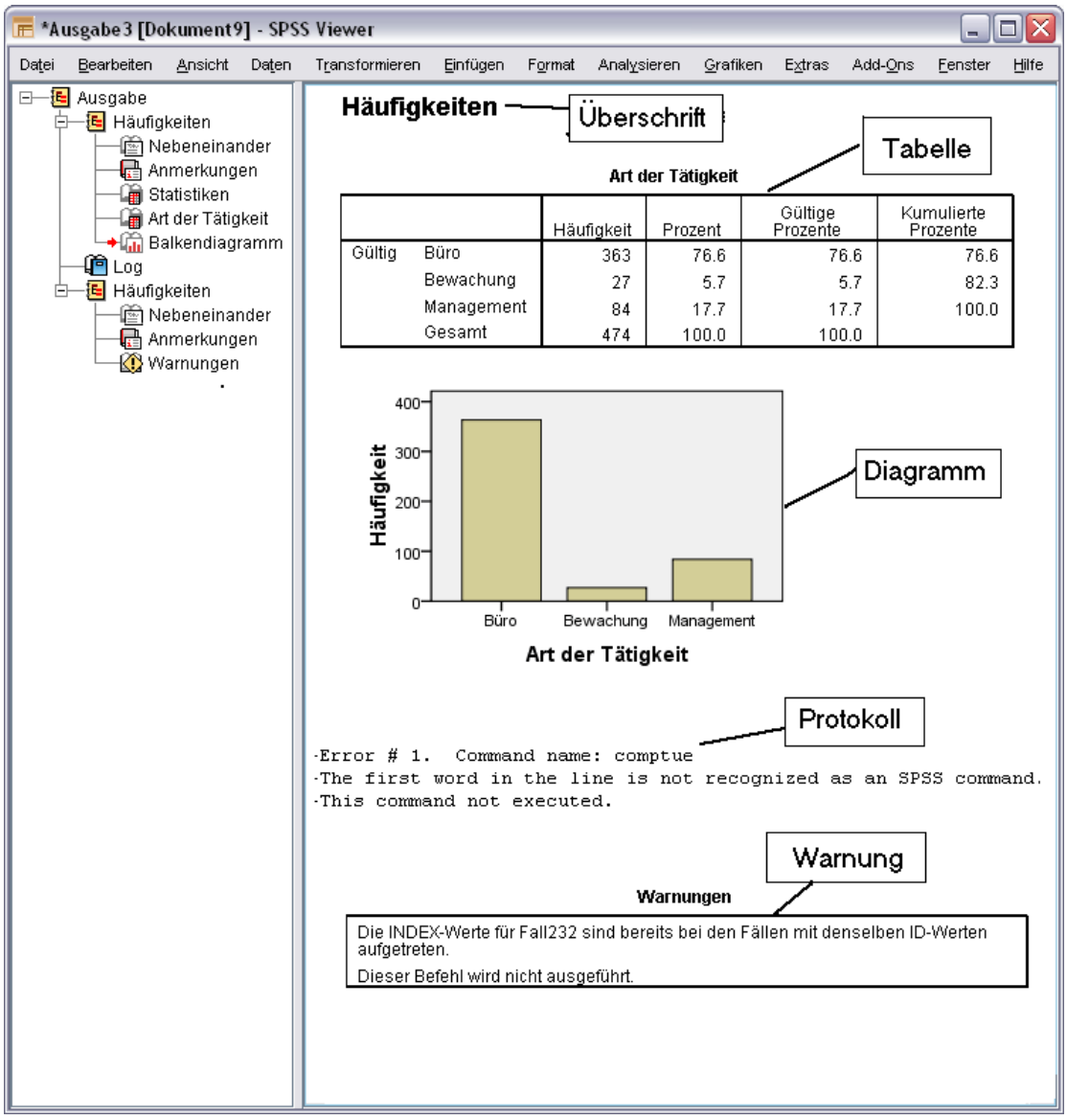
Texte. Textobjekte, die weder Logs noch Überschriften sind (einschließlich der Objekte, die im Gliederungsfenster des Viewers mit *Textausgabe* gekennzeichnet sind).

Überschriften. Textobjekte, die im Gliederungsfenster des Viewers mit *Titel* gekennzeichnet sind. Überschriften-Textobjekte werden beim Ausgabe-XML-Format nicht berücksichtigt.

Warnungen. Warnungensammlungen enthalten bestimmte Arten von Fehler- und Warnmeldungen.

Bäume. Baummodelldiagramme, die mit der Option "Classification Tree" erzeugt wurden. Baumobjekte werden nur bei den Zielformaten XML und HTML berücksichtigt.

Abbildung 48-2
Ausgabeobjekttypen



Befehls-IDs und Tabellenuntertypen

Befehls-IDs

Befehls-IDs sind für alle Statistik- und Diagrammprozeduren verfügbar, außerdem für alle Befehle, bei denen Ausgabeblöcke mit eigener identifizierbarer Überschrift im Gliederungsfenster des Viewers erzeugt werden. Diese IDs sind in der Regel (jedoch nicht immer) identisch oder nahezu identisch mit den Namen der Prozeduren in den Menüs und den Dialogfeldtiteln, die wiederum in der Regel (jedoch nicht immer) identisch oder nahezu identisch mit den Bezeichnungen der

zugrunde liegenden Befehlsnamen sind. Die Befehls-ID für die Prozedur “Häufigkeiten” lautet beispielsweise “Häufigkeiten”; dies ist auch die Bezeichnung des zugrunde liegenden Befehls.

In einigen Fällen weisen der Name der Prozedur und die Befehls-ID und/oder der Befehlsname allerdings beträchtliche Unterschiede auf. Beispielsweise greifen alle Prozeduren im Untermenü “Nichtparametrisch” des Menüs “Analysieren” auf denselben zugrunde liegenden Befehl zurück, und die Befehls-ID ist mit dem Namen des zugrunde liegenden Befehls identisch: Npar Tests.

Tabellenuntertypen

Tabellenuntertypen sind die verschiedenen Typen von Pivot-Tabellen, die erstellt werden können. Einige Untertypen werden nur von einem einzigen Befehl erzeugt, andere Untertypen dagegen von mehreren Befehlen (die Tabellen zeigen jedoch unter Umständen nicht dasselbe Erscheinungsbild). Die Namen der Tabellenuntertypen sind normalerweise aussagekräftig. Es können allerdings zahlreiche Untertypen zur Auswahl stehen (insbesondere wenn Sie viele Befehle ausgewählt haben). Zwei Untertypen können auch sehr ähnliche Namen besitzen.

So suchen Sie Befehls-IDs und Tabellenuntertypen:

Im Zweifelsfall können Sie die Befehls-IDs und die Namen der Tabellenuntertypen im Viewer-Fenster suchen:

- Starten Sie die Prozedur, um Ausgaben im Viewer zu erzeugen.
- Klicken Sie mit der rechten Maustaste auf das Element im Gliederungsfenster des Viewers.
- Wählen Sie die Option OMS-Befehls-ID kopieren oder die Option OMS-Tabellenuntertyp kopieren.
- Fügen Sie die kopierte Befehls-ID bzw. den Namen des Untertabellentyps in einen Text-Editor ein (z. B. in ein Syntax-Editor-Fenster).

Labels

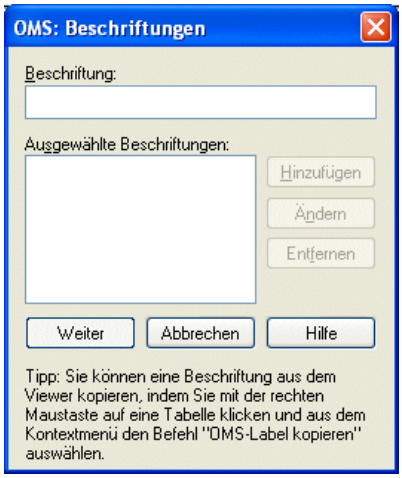
Als Alternative zu Namen von Tabellenuntertypen können Sie Tabellen auf der Grundlage des Texts auswählen, der im Gliederungsfenster des Viewers angezeigt wird. Sie können auch andere Objekttypen anhand der Beschriftung auswählen. Beschriftungen helfen beim Unterscheiden zwischen mehreren Tabellen desselben Typs, bei denen der Gliederungstext ein Attribut des jeweiligen Ausgabeobjekts angibt, beispielsweise die Variablennamen oder -beschriftungen. Es gibt jedoch eine Reihe von Faktoren, die den Labeltext beeinflussen können:

- Wenn die Verarbeitung aufgeteilter Dateien aktiviert ist, kann die Gruppen-ID für die aufgeteilte Datei an die Beschriftung angehängt werden.
- Beschriftungen mit Informationen zu Variablen oder Werten sind abhängig von den aktuellen Einstellungen für die Beschriftung der Ausgabe (Menü “Bearbeiten”, “Optionen”, Registerkarte “Beschriftung der Ausgabe”).
- Labels richten sich außerdem nach der aktuellen Einstellung für die Ausgabesprache (Menü “Bearbeiten”, “Optionen”, Registerkarte “Allgemein”).

So legen Sie Beschriftungen zum Identifizieren von Ausgabetabellen fest:

- Wählen Sie im Bedienfeld des Ausgabeverwaltungssystems mindestens einen Ausgabetyp und anschließend mindestens einen Befehl aus.
- Klicken Sie auf Beschriftungen.

Abbildung 48-3
Dialogfeld "OMS: Beschriftungen"



- Geben Sie die Beschriftung *auf dieselbe Weise* ein, wie sie im Gliederungsfenster des Viewer-Fensters aufgeführt wird. (Alternativ können Sie mit der rechten Maustaste auf das Element in der Gliederung klicken, die Option OMS-Label kopieren auswählen und die kopierte Beschriftung dann im Textfeld "Beschriftung" einfügen.)
- Klicken Sie auf Hinzufügen.
- Wiederholen Sie diesen Vorgang für jede Beschriftung, die Sie hinzufügen möchten.
- Klicken Sie auf Weiter.

Platzhalter

Sie können ein Sternchen (*) als Platzhalterzeichen als letztes Zeichen im Bezeichnungs-String verwenden. Alle Beschriftungen, die mit dem angegebenen String beginnen (alle Zeichen mit Ausnahme des Sternchens), werden ausgewählt. Dies ist nur dann möglich, wenn das Sternchen das letzte Zeichen ist, weil Sternchen durchaus als zulässige Zeichen innerhalb einer Beschriftung auftreten können.

OMS: Optionen

Das Dialogfeld "OMS: Optionen" bietet die folgenden Möglichkeiten:

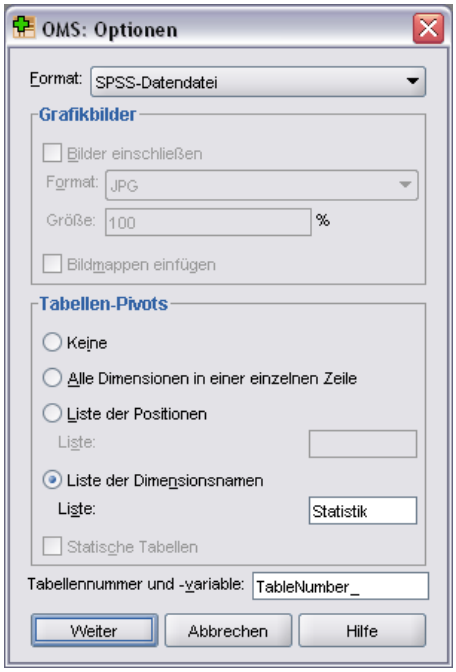
- Legen Sie das Ausgabeformat fest.
- Geben Sie das Bildformat an (bei den Ausgabeformaten "HTML" und "Ausgabe-XML").

- Geben Sie an, welche Tabellendimensionselemente in die Zeilendimension eingehen sollen.
- Schließen Sie eine Variable für die laufende Tabellennummer ein, die als Quelle für die einzelnen Fälle herangezogen werden soll (beim SPSS-Datendateiformat).

So legen Sie OMS-Optionen fest:

▶ Klicken Siein der Systemsteuerung des Ausgabeverwaltungssystems (OMS) auf Optionen.

Abbildung 48-4
Dialogfeld "OMS: Optionen"



Format

Ausgabe-XML. XML, das dem Schema *spss-output* entspricht. Diagramme werden als XML berücksichtigt, das dem Schema "vizml" entspricht (*xml.spss.com/spss/visualization*). Baumdiagramme werden als XML berücksichtigt, das dem Schema "pmml" entspricht (*www.dmg.org*). Anstatt Diagramme und Baumdiagramme im XML-Format aufzunehmen, können Sie sie als sparate Dateien im ausgewählten Grafikformat speichern.

HTML. Ausgabeobjekte, die als Pivot-Tabellen im Viewer dargestellt würden, werden in einfache HTML-Tabellen umgewandelt. Attribute aus Tabellenvorlagen (z. B. Schriftart, Rahmenformate, Farben) werden nicht unterstützt. Textausgabeobjekte werden als <PRE> in der HTML gekennzeichnet. Wenn Sie die Diagramme mit aufnehmen, werden diese als separate Dateien im ausgewählten Grafikformat exportiert und als Verweis eingebettet.

SPSS-Datendatei. Hierbei handelt es sich um ein binäres Dateiformat. Alle Ausgabeobjekttypen mit Ausnahme der Tabellen sind ausgeschlossen. Jede Spalte einer Tabelle wird zu einer Variablen in der Datendatei. Soll eine mit OMS erstellte Datendatei noch in derselben Sitzung genutzt werden, beenden Sie die aktive OMS-Anforderung, damit Sie die Datendatei öffnen können. Für weitere Informationen siehe Weiterleiten der Ausgabe an SPSS-Datendateien auf S. 535.

SPV. SPSS-Viewer-Dateiformat Es handelt sich hierbei um dasselbe Format wie beim Speichern der Inhalte eines Viewer-Fensters.

Text. Text, der mit Leerzeichen getrennt ist. Die Ausgabe wird als Text geschrieben. Bei Schriftarten mit fester Breite wird die Tabellenausgabe mithilfe von Leerzeichen ausgerichtet. Alle Diagramme und werden ausgeschlossen.

Tabulatorgetrennter Text. Text, der mit Tabulatoren getrennt ist. Bei Ausgaben, die im Viewer als Pivot-Tabellen angezeigt werden, begrenzen die Tabulatoren die Tabellenspaltenelemente. Textblockzeilen werden unverändert geschrieben; der Text wird nicht mit Tabulatoren an sinnvollen Positionen gegliedert. Alle Diagramme und werden ausgeschlossen.

Grafiken

Bei den Formaten "HTML" und "Ausgabe-XML" können Sie Diagramme und Baummodelldiagramme als Bilddateien einschließen. Für jedes Diagramm bzw. jeden Baum wird eine eigene Bilddatei erstellt.

- Beim Dokumentformat "HTML" werden für jede Bilddatei standardmäßige Tags vom Typ `<IMG SRC='dateiname'>` in das HMTL-Dokument aufgenommen.
- Beim Dokumentformat "Ausgabe-XML" enthält die XML-Datei für jede Bilddatei ein `chart`-Element mit einem `ImageFile`-Attribut im allgemeinen Format `<chart imageFile="dateipfad/dateiname"/>`.
- Die Bilddateien werden in einem separaten Unterverzeichnis (Ordner) gespeichert. Der Name des Unterverzeichnisses ist der Name der Zieldatei, ohne Dateinamenerweiterung, dafür mit dem Suffix *_files*. Wenn die Zieldatei beispielsweise den Namen *julidaten.htm* trägt, erhält das Bildunterverzeichnis den Namen *julidaten_files*.

Format. Als Bildformate stehen PNG, JPG, EMF, BMP und VML zur Verfügung.

- Das Format EMF (Enhanced Metafile) ist nur unter Windows-Betriebssystemen verfügbar.
- Das Bildformat "VML" ist nur für das Dokumentformat "HTML" verfügbar.
- Beim Bildformat "VML" werden keine separaten Bilddateien erstellt. Der VML-Code, der das Bild rendert, ist im HTML-Code eingebettet.
- Das Bildformat "VML" beinhaltete keine Baumdiagramme.

Größe. Sie können das Bild von 10 % bis 200 % skalieren.

Bildmappen einfügen. Beim Dokumentformat "HTML" erstellt diese Option QuickInfos für Bildmappen (Image Maps), die Informationen für bestimmte Diagrammelemente anzeigen, beispielsweise den Wert des ausgewählten Punkts in einem Liniendiagramm bzw. des ausgewählten Balkens in einem Balkendiagramm.

Tabellen-Pivots

Bei der Ausgabe von Pivot-Tabellen können Sie das oder die Dimensionselemente bestimmen, die in den Spalten auftreten sollen. Alle anderen Dimensionselemente treten in den Zeilen auf. Beim SPSS-Datendateiformat werden die Tabellenspalten zu Variablen und die Zeilen zu Fällen.

- Wenn Sie mehrere Dimensionselemente für die Spalten angeben, werden diese Elemente in der Reihenfolge in den Spalten verschachtelt, in der sie aufgeführt sind. Beim SPSS-Datendateiformat werden die Variablennamen aus verschachtelten Spaltenelementen gebildet. Für weitere Informationen siehe Variablennamen in Datendateien aus dem OMS auf S. 542.
- Wenn eine Tabelle keine der aufgeführten Dimensionselemente enthält, werden alle Dimensionselemente dieser Tabelle in den Zeilen aufgeführt.
- Die hier angegebenen Tabellen-Pivots wirken sich nicht auf die Tabellen aus, die im Viewer dargestellt werden.

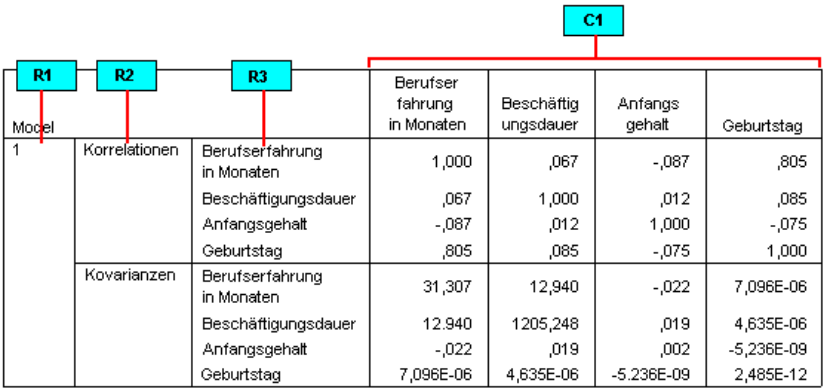
Jede Dimension einer Tabelle (Zeile, Spalte, Schicht) kann null oder mehr Elemente enthalten. Eine einfache Kreuztabelle mit zwei Dimensionen enthält beispielsweise ein einziges Zeilendimensionselement und ein einziges Spaltendimensionselement, die jeweils eine der in der Tabelle verwendeten Variablen enthalten. Die Dimensionselemente für die Spaltendimension können wahlweise mithilfe von Positionsargumenten oder mit den Dimensionselement-"Namen" festgelegt werden.

Alle Dimensionen in einer einzelnen Zeile. Hiermit wird eine einzelne Zeile für jede Tabelle erstellt. Bei Datendateien im SPSS-Format bedeutet dies, dass jede Tabelle einen einzelnen Fall darstellt und alle Tabellenelemente Variablen sind.

Liste der Positionen. Ein Positionsargument besteht in der Regel aus einem Buchstaben für die Standardposition des Elements (C für Spalte, R für Zeile, L für Schicht), gefolgt von einer positiven ganze Zahl, aus der die Standardposition innerhalb dieser Dimension hervorgeht. R1 bezeichnet beispielsweise das äußerste Zeilendimensionselement.

- Sollen mehrere Elemente aus mehreren Dimensionen angegeben werden, trennen Sie die einzelnen Dimensionen jeweils mit einem Leerzeichen, z. B. R1 C2.
- Steht nach dem Dimensionsbuchstaben die Zeichenfolge "ALL", bedeutet dies, dass alle Elemente in der betreffenden Dimension in ihrer Standardreihenfolge berücksichtigt werden. CALL entspricht beispielsweise dem Standardverhalten; die Spalten werden hierbei auf der Grundlage aller Spaltenelemente in ihrer Standardreihenfolge gebildet.
- Mit CALL RALL LALL (oder RALL CALL LALL usw.) werden alle Dimensionselemente in die Spalten aufgenommen. Beim SPSS-Datendateiformat entsteht hierbei je eine Zeile/ein Fall pro Tabelle in der Datendatei.

Abbildung 48-5
Zeilen- und Spaltenpositionsargumente



Modell			Berufserfahrung in Monaten	Beschäftigungsdauer	Anfangsgehalt	Geburtstag
1	Korrelationen	Berufserfahrung in Monaten	1,000	,067	-,087	,805
		Beschäftigungsdauer	,067	1,000	,012	,085
		Anfangsgehalt	-,087	,012	1,000	-,075
		Geburtstag	,805	,085	-,075	1,000
	Kovarianzen	Berufserfahrung in Monaten	31,307	12,940	-,022	7,096E-06
		Beschäftigungsdauer	12.940	1205,248	,019	4,635E-06
		Anfangsgehalt	-,022	,019	,002	-5,236E-09
		Geburtstag	7,096E-06	4,635E-06	-5.236E-09	2,485E-12

Liste der Dimensionsnamen. Als Alternative zu Positionsargumenten können Sie die "Namen" der Dimensionselemente verwenden, also die Textbeschriftungen, die in der Tabelle aufgeführt werden. Eine einfache zweidimensionale Kreuztabelle enthält beispielsweise ein einziges Zeilendimensionselement und ein einziges Spaltendimensionselement, die jeweils mit einer Beschriftung auf der Grundlage der Variablen in diesen Dimensionen versehen sind, außerdem ein einziges Schichtdimensionselement mit der Beschriftung *Statistik* (wenn Deutsch die Ausgabesprache ist).

- Die Dimensionselementnamen sind abhängig von der Ausgabesprache und/oder von den Einstellungen, die sich auf die Anzeige von Variablennamen und/oder Beschriftungen in Tabellen auswirken.
- Jeder Dimensionselementname muss in einfache oder doppelte Anführungszeichen eingeschlossen werden. Sollen mehrere Dimensionselementnamen angegeben werden, trennen Sie die einzelnen, in Anführungsstrichen stehenden Namen jeweils mit einem Leerzeichen.

Die Beschriftungen für die Dimensionselemente sind nicht in jedem Fall deutlich.

So lassen Sie alle Dimensionselemente und deren Beschriftungen für eine Pivot-Tabelle anzeigen:

▶ Aktivieren Sie die Tabelle im Viewer durch Doppelklicken.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Ansicht
  Alles einblenden

und/oder

▶ Falls die Pivot-Leisten nicht angezeigt werden, wählen Sie folgende Befehle aus den Menüs aus:

Pivot
  Pivot-Leisten

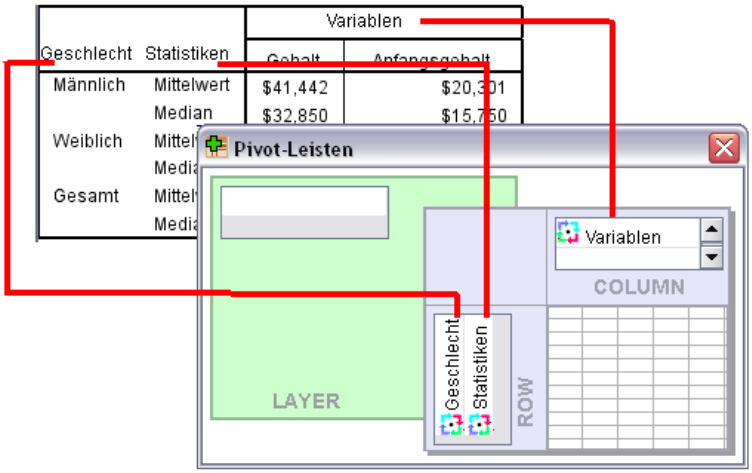
Die Elementbeschriftungen werden in den Pivot-Leisten angezeigt.

Abbildung 48-6
Dimensionselementnamen in der Tabelle und in den Pivot-Leisten

Tabelle ohne Anzeige der Beschriftungen für die Dimensionselemente

		Gehalt	Anfangsgehalt
Männlich	Mittelwert	$41,442	$20,301
	Median	$32,850	$15,750
Weiblich	Mittelwert	$26,032	$13,092
	Median	$24,300	$12,375
Gesamt	Mittelwert	$34,420	$17,016
	Median	$28,875	$15,000

Tabelle mit Anzeige der Beschriftungen für die Dimensionselemente



Protokollierung

Sie können die OMS-Aktivitäten in einem Protokoll im XML- oder Textformat protokollieren lassen.

- Im Protokoll werden alle neuen OMS-Anforderungen für die aktuelle Sitzung aufgezeichnet, nicht jedoch OMS-Anforderungen, die bereits aktiv waren, bevor Sie die Protokollierung aktiviert haben.
- Die aktuelle Protokolldatei wird beendet, sobald Sie eine neue Protokolldatei angeben oder die Option OMS-Aktivität protokollieren deaktivieren.

So aktivieren Sie die OMS-Protokollierung:

▶ Klicken Siein der Systemsteuerung des Ausgabeverwaltungssystems (OMS) auf Protokollierung.

Ausschließen der Ausgabeanzeige aus dem Viewer

Das Kontrollkästchen Aus Viewer ausschließen unterdrückt die Anzeige aller Ausgaben im Viewer-Fenster, die in der OMS-Anforderung ausgewählt wurden. Dies eignet sich insbesondere für Produktionsjobs, bei denen eine umfangreiche Ausgabe entsteht, ohne dass die Ergebnisse in einem Viewer-Dokument (.spv-Datei) dargestellt werden müssen. Darüber hinaus können Sie

mit dieser Funktion die Anzeige bestimmter Ausgabeobjekte unterdrücken, die einfach nicht dargestellt werden sollen, ohne andere Ausgaben an eine externe Datei und in einem anderen Format weiterzuleiten.

So unterdrücken Sie die Anzeige bestimmter Ausgabeobjekte, ohne andere Ausgaben an eine externe Datei weiterzuleiten:

- Erstellen Sie eine OMS-Anforderung, mit der die unerwünschte Ausgabe ermittelt wird.
- Wählen Sie Aus Viewer ausschließen.
- Wählen Sie für das Ausgabeziel Datei, aber machen Sie keine Dateiangabe.
- Klicken Sie auf Hinzufügen.

Die ausgewählte Ausgabe wird aus dem Viewer ausgeschlossen; die restliche Ausgabe wird wie gewohnt im Viewer dargestellt.

Weiterleiten der Ausgabe an SPSS-Datendateien

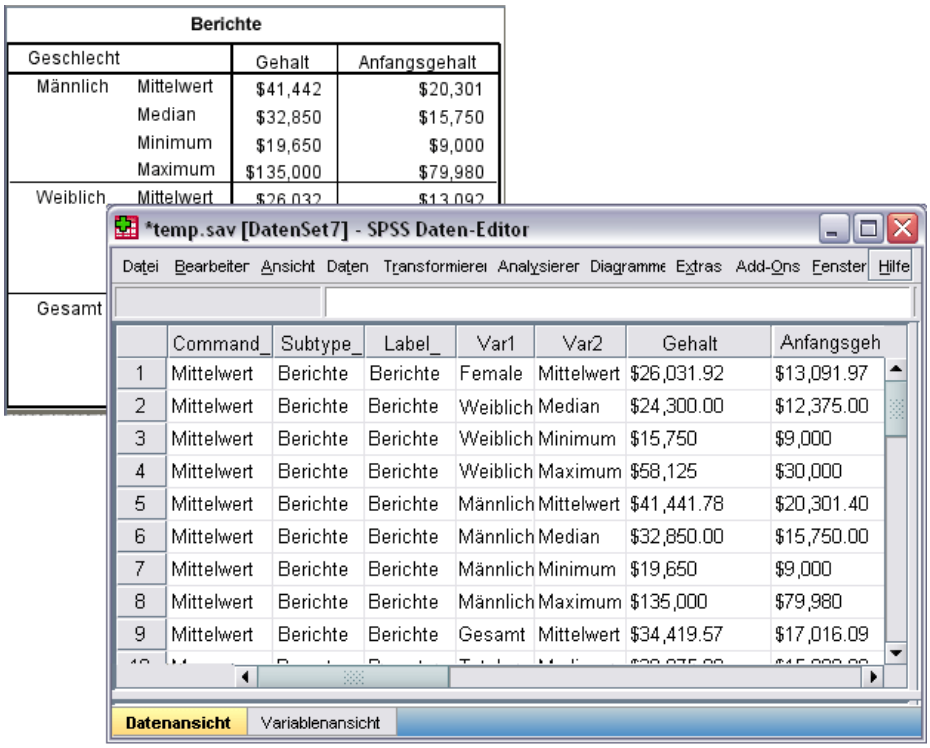
Eine SPSS-Datendatei besteht aus Variablen (in den Spalten) und Fällen (in den Zeilen). Im Wesentlichen entspricht dies dem Verfahren, wie Pivot-Tabellen in Datendateien umgewandelt werden:

- Spalten in der Tabelle werden zu Variablen in der Datendatei. Aus den Spaltenbeschriftungen werden gültige Variablennamen gebildet.
- Die Zeilenbeschriftungen in der Spalte werden zu Variablen mit generischen Variablennamen (*Var1*, *Var2*, *Var3*) in der Datendatei. Die Werte dieser Variablen entsprechen den Zeilenbeschriftungen in der Tabelle.
- In die Datendatei werden automatisch drei Tabellen-ID-Variablen aufgenommen: *Command_*, *Subtype_* und *Label_*. Alle drei Variablen sind String-Variablen. Die ersten beiden Variablen bezeichnen die Befehls- und die Untertyp-ID. Für weitere Informationen siehe Befehls-IDs und Tabellenuntertypen auf S. 527. *Label_* enthält den Tabellentiteltext.
- Zeilen in der Tabelle werden zu Fällen in der Datendatei.

Beispiel: Einzelne zweidimensionale Tabelle

Im einfachsten Fall (also bei einer einzelnen, zweidimensionalen Tabelle) werden die Tabellenspalten zu Variablen und die Zeilen zu Fällen in der Datendatei.

Abbildung 48-7
Einzelne zweidimensionale Tabelle



Berichte

Geschlecht		Gehalt	Anfangsgehalt
Männlich	Mittelwert	$41,442	$20,301
	Median	$32,850	$15,750
	Minimum	$19,650	$9,000
	Maximum	$135,000	$79,980
Weiblich	Mittelwert	$26,032	$13,092
Gesamt			

*temp.sav [DatenSet7] - SPSS Daten-Editor

Datei Bearbeiter Ansicht Daten Transformieren Analysieren Diagramme Extras Add-Ons Fenster Hilfe

	Command_	Subtype_	Label_	Var1	Var2	Gehalt	Anfangsgeh
1	Mittelwert	Berichte	Berichte	Female	Mittelwert	$26,031.92	$13,091.97
2	Mittelwert	Berichte	Berichte	Weiblich	Median	$24,300.00	$12,375.00
3	Mittelwert	Berichte	Berichte	Weiblich	Minimum	$15,750	$9,000
4	Mittelwert	Berichte	Berichte	Weiblich	Maximum	$58,125	$30,000
5	Mittelwert	Berichte	Berichte	Männlich	Mittelwert	$41,441.78	$20,301.40
6	Mittelwert	Berichte	Berichte	Männlich	Median	$32,850.00	$15,750.00
7	Mittelwert	Berichte	Berichte	Männlich	Minimum	$19,650	$9,000
8	Mittelwert	Berichte	Berichte	Männlich	Maximum	$135,000	$79,980
9	Mittelwert	Berichte	Berichte	Gesamt	Mittelwert	$34,419.57	$17,016.09

Datenansicht Variablenansicht

- Die ersten drei Variablen kennzeichnen die Quelltabelle anhand des Befehls, des Untertyps und der Beschriftung.
- Die beiden Elemente, die die Zeilen in der Tabelle definiert hatten (Werte für die Variable *Geschlecht* und statistische Maße) werden mit den generischen Variablennamen *Var1* und *Var2* versehen. Beide Variablen sind String-Variablen.
- Aus den Spaltenbeschriftungen in der Tabelle werden gültige Variablennamen gebildet. In diesem Fall beruhen diese Variablennamen auf den Variablenlabels der drei in der Tabelle ausgewerteten metrischen Variablen. Falls für die Variablen keine Variablenlabels definiert sind oder die Variablennamen anstelle der Variablenlabels als Spaltenbeschriftungen in der Tabelle angezeigt werden sollen, wären die Variablennamen in der neuen Datendatei mit den Namen in der Quelldatendatei identisch.

Beispiel: Tabellen mit Schichten

Neben Zeilen und Spalten kann eine Tabelle eine dritte Dimension aufweisen: die Schichtdimension.

Abbildung 48-8
Tabelle mit Schichten

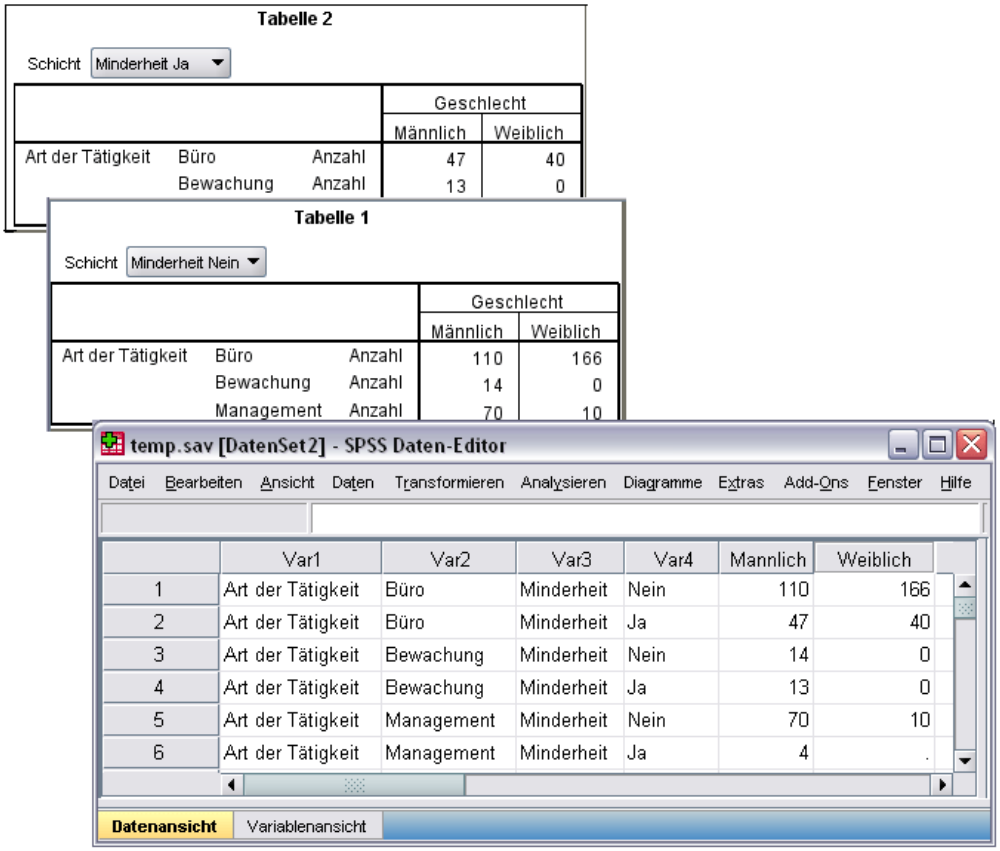


Tabelle 2

Schicht: Minderheit Ja

			Geschlecht	
			Männlich	Weiblich
Art der Tätigkeit	Büro	Anzahl	47	40
	Bewachung	Anzahl	13	0

Tabelle 1

Schicht: Minderheit Nein

			Geschlecht	
			Männlich	Weiblich
Art der Tätigkeit	Büro	Anzahl	110	166
	Bewachung	Anzahl	14	0
	Management	Anzahl	70	10

temp.sav [DatenSet2] - SPSS Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Diagramme Extras Add-Ons Fenster Hilfe

	Var1	Var2	Var3	Var4	Mannlich	Weiblich
1	Art der Tätigkeit	Büro	Minderheit	Nein	110	166
2	Art der Tätigkeit	Büro	Minderheit	Ja	47	40
3	Art der Tätigkeit	Bewachung	Minderheit	Nein	14	0
4	Art der Tätigkeit	Bewachung	Minderheit	Ja	13	0
5	Art der Tätigkeit	Management	Minderheit	Nein	70	10
6	Art der Tätigkeit	Management	Minderheit	Ja	4	.

Datenansicht | Variablenansicht

- In der Tabelle werden die Schichten durch die Variable *Minderheit* definiert. In der Datendatei werden so zwei zusätzliche Variablen erstellt: eine Variable, die das Schichtelement identifiziert, und eine Variable, die die Kategorien des Schichtelements bezeichnet.
- Die Variablen, die aus den Schichtelementen gebildet wurden, sind ebenfalls String-Variablen mit generischen Variablennamen (Präfix *Var*, gefolgt von einer laufenden Nummer), so wie die Variablen aus den Zeilenelementen.

Datendateien aus mehreren Tabellen

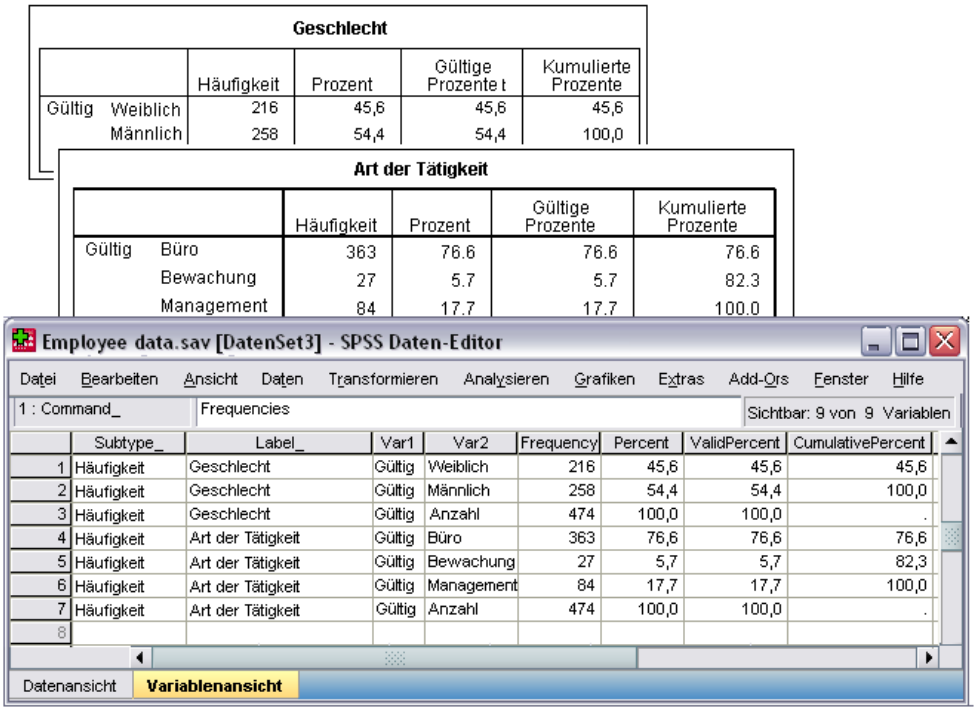
Werden mehrere Tabellen an dieselbe Datendatei weitergeleitet, wird jede Tabelle jeweils zur Datendatei hinzugefügt. Ähnlich wie beim Zusammenfügen von Datendateien werden hierbei die Fälle aus einer Datendatei in eine andere Datendatei aufgenommen (Menü “Daten”, “Dateien zusammenfügen”, “Fälle hinzufügen”).

- Bei jeder nachfolgenden Tabelle werden weitere Fälle zur Datendatei hinzugefügt.
- Weichen die Spaltenbeschriftungen in den Tabellen voneinander ab, können ggf. auch Variablen in die Datendatei aufgenommen werden. Bei Fällen aus anderen Tabellen, die keine Spalte mit der entsprechenden Beschriftung aufweisen, entstehen dabei fehlende Werte.

Beispiel: Mehrere Tabellen mit identischen Spaltenbeschriftungen

Mehrere Tabellen, die dieselben Spaltenbeschriftungen enthalten, resultieren in der Regel in sofort nutzbaren Datendateien; diese Dateien müssen nicht mehr nachbearbeitet werden. Die mit der Prozedur “Häufigkeit” erzeugten Häufigkeitstabellen besitzen beispielsweise immer dieselben Spaltenbeschriftungen.

Abbildung 48-9
Zwei Tabellen mit identischen Spaltenbeschriftungen



Geschlecht

		Häufigkeit	Prozent	Gültige Prozente t	Kumulierte Prozente
Gültig	Weiblich	216	45,6	45,6	45,6
	Männlich	258	54,4	54,4	100,0

Art der Tätigkeit

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Büro	363	76,6	76,6	76,6
	Bewachung	27	5,7	5,7	82,3
	Management	84	17,7	17,7	100,0

Employee data.sav [DatenSet3] - SPSS Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Grafiken Extras Add-Ons Fenster Hilfe

1 : Command_ | Frequencies | Sichtbar: 9 von 9 Variablen

	Subtype_	Label_	Var1	Var2	Frequency	Percent	ValidPercent	CumulativePercent
1	Häufigkeit	Geschlecht	Gültig	Weiblich	216	45,6	45,6	45,6
2	Häufigkeit	Geschlecht	Gültig	Männlich	258	54,4	54,4	100,0
3	Häufigkeit	Geschlecht	Gültig	Anzahl	474	100,0	100,0	.
4	Häufigkeit	Art der Tätigkeit	Gültig	Büro	363	76,6	76,6	76,6
5	Häufigkeit	Art der Tätigkeit	Gültig	Bewachung	27	5,7	5,7	82,3
6	Häufigkeit	Art der Tätigkeit	Gültig	Management	84	17,7	17,7	100,0
7	Häufigkeit	Art der Tätigkeit	Gültig	Anzahl	474	100,0	100,0	.
8								

Datenansicht | Variablenansicht

- Die zweite Tabelle trägt weitere Fälle (Zeilen) zur Datendatei bei, führt jedoch keine neuen Variablen ein, weil die Spaltenbeschriftungen exakt übereinstimmen. So entstehen keine großen Bereiche mit fehlenden Daten.
- Die Werte für *Command_* und *Subtype_*+ sind identisch; der Wert für *Label_* zeigt die Quelltabelle für die einzelnen Fallgruppen, weil die beiden Häufigkeitstabellen einen anderen Titel besitzen.

Beispiel: Mehrere Tabellen mit verschiedenen Spaltenbeschriftungen

Für jede eindeutige Spaltenbeschriftung in den Tabellen, die an die Datendatei weitergeleitet werden, wird eine neue Variable in der Datendatei erstellt. Wenn die Tabellen unterschiedliche Spaltenbeschriftungen enthalten, führt dies zu Bereichen mit fehlenden Daten.

Abbildung 48-10
Zwei Tabellen mit unterschiedlichen Spaltenbeschriftungen

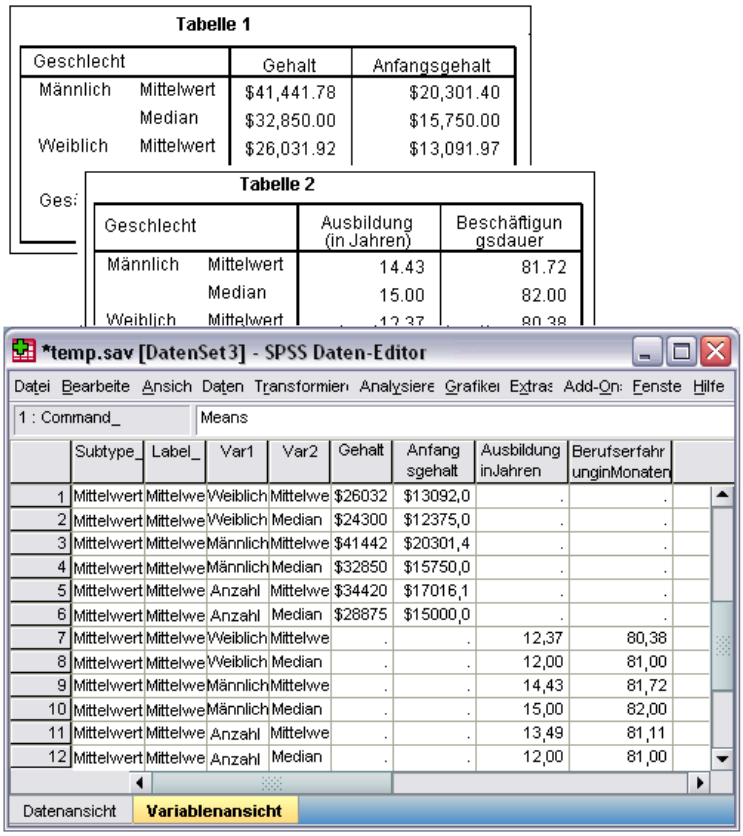


Tabelle 1

Geschlecht		Gehalt	Anfangsgehalt
Männlich	Mittelwert	$41,441.78	$20,301.40
	Median	$32,850.00	$15,750.00
Weiblich	Mittelwert	$26,031.92	$13,091.97

Tabelle 2

Geschlecht		Ausbildung (in Jahren)	Beschäftigungsdauer
Männlich	Mittelwert	14.43	81.72
	Median	15.00	82.00
Weiblich	Mittelwert	12.37	80.38

*temp.sav [DatenSet3] - SPSS Daten-Editor

Datei Bearbeite Ansich Daten Transformier Analysiere Grafiker Extras Add-On: Fenste Hilfe

1 : Command_ | Means

	Subtype_	Label_	Var1	Var2	Gehalt	Anfangsgehalt	Ausbildung inJahren	Berufserfahrung inMonaten
1	Mittelwert	Mittelwe	Weiblich	Mittelwe	$26032	$13092,0	.	.
2	Mittelwert	Mittelwe	Weiblich	Median	$24300	$12375,0	.	.
3	Mittelwert	Mittelwe	Männlich	Mittelwe	$41442	$20301,4	.	.
4	Mittelwert	Mittelwe	Männlich	Median	$32850	$15750,0	.	.
5	Mittelwert	Mittelwe	Anzahl	Mittelwe	$34420	$17016,1	.	.
6	Mittelwert	Mittelwe	Anzahl	Median	$28875	$15000,0	.	.
7	Mittelwert	Mittelwe	Weiblich	Mittelwe	.	.	12,37	80,38
8	Mittelwert	Mittelwe	Weiblich	Median	.	.	12,00	81,00
9	Mittelwert	Mittelwe	Männlich	Mittelwe	.	.	14,43	81,72
10	Mittelwert	Mittelwe	Männlich	Median	.	.	15,00	82,00
11	Mittelwert	Mittelwe	Anzahl	Mittelwe	.	.	13,49	81,11
12	Mittelwert	Mittelwe	Anzahl	Median	.	.	12,00	81,00

Datenansicht | Variablenansicht

- Die erste Tabelle enthält die Spalten *Anfangsgehalt* und *Aktuelles Gehalt*, die in der zweiten Tabelle nicht vorhanden sind. Bei Fällen aus der zweiten Tabelle entstehen so fehlende Werte für diese Variablen.
- Umgekehrt enthält die zweite Tabelle die Spalten *Schulabschluss* und *Beschäftigungsdauer*, die in der ersten Tabelle nicht vorhanden sind. Bei Fällen aus der ersten Tabelle entstehen entsprechend fehlende Werte für diese Variablen.
- Nicht übereinstimmende Variablen wie in diesem Beispiel können sogar bei Tabellen auftreten, die denselben Untertyp aufweisen. In diesem Beispiel besitzen beide Tabellen denselben Untertyp.

Beispiel: Datendateien, die nicht aus mehreren Tabellen erzeugt wurden

Falls eine Tabelle nicht dieselbe Anzahl an Zeilenelementen enthält wie die anderen Tabellen, wird keine Datendatei erstellt. Hierbei muss nicht die Anzahl der Zeilen identisch sein, sondern die Anzahl der Zeilen*elemente*, die zu Variablen in der Datendatei werden. Beispiel: Eine Kreuztabelle mit zwei Variablen und eine Kreuztabelle mit drei Variablen enthalten verschieden viele Zeilenelemente, weil die Variable “layer” bei der Standard-Kreuztabellenanzeige mit drei Variablen in der Zeilenvariable verschachtelt ist.

Abbildung 48-11
Tabellen mit unterschiedlich vielen Zeilenelementen

Art der Tätigkeit * Geschlecht Kreuztabelle

		Geschlecht		Gesamt
		Männlich	Weiblich	
Art der Tätigkeit	Büro	157	206	363
	Bewachung	27	0	27
	Management	74	10	84
Gesamt		258	216	474

Art der Tätigkeit * Geschlecht * Minderheit Kreuztabelle

			Geschlecht		Gesamt
Minderheit			Männlich	Weiblich	
Nein	Art der Tätigkeit	Büro	110	166	276
		Bewachung	14	0	14
		Management	70	10	80
	Gesamt		194	176	370
Ja	Art der Tätigkeit	Büro	47	40	87
		Bewachung	13	0	13
		Management	4	0	4
	Gesamt		64	40	104

Steuern von Spaltenelementen zum Steuern von Variablen in der Datendatei

Im Dialogfeld “Optionen” des Ausgabeverwaltungssystem-Bedienfelds können Sie angeben, welche Dimensionselemente sich in den Spalten befinden und daher zum Erstellen der Variablen in der erzeugten Datendatei verwendet werden sollen. Dies entspricht dem Pivotieren der Tabelle im Viewer.

Beispiel: Bei der Prozedur “Häufigkeiten” wird eine deskriptive Statistiktabelle mit Statistiken in den Zeilen gebildet, bei der Prozedur “Deskriptive Statistiken” dagegen eine deskriptive Statistiktabelle mit Statistiken in den Zeilen. Wenn beide Tabellentypen sinnvoll in dieselbe Datendatei aufgenommen werden sollen, muss die Spaltendimension für eine der Tabellentypen geändert werden.

Bei beiden Tabellentypen trägt die Statistikdimension den Elementnamen “Statistik”. Um die statistischen Werte aus der Häufigkeitstabelle in die Spalten zu übernehmen, reicht es daher, den Eintrag “Statistik” (in Anführungszeichen) zur Liste der Dimensionsnamen im Dialogfeld “Optionen” hinzuzufügen.

Abbildung 48-12
Dialogfeld "OMS: Optionen"

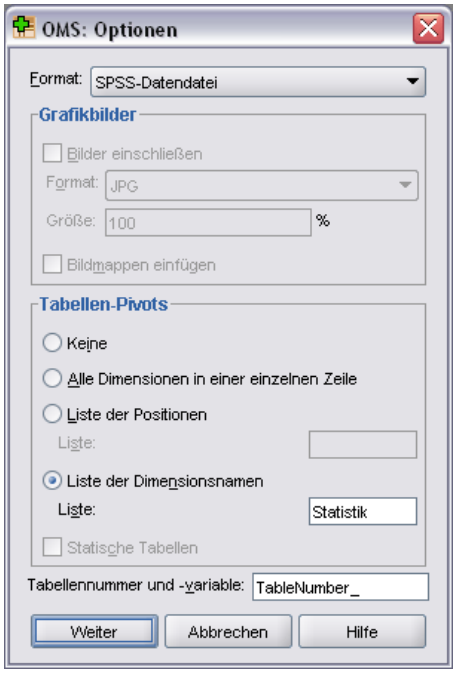
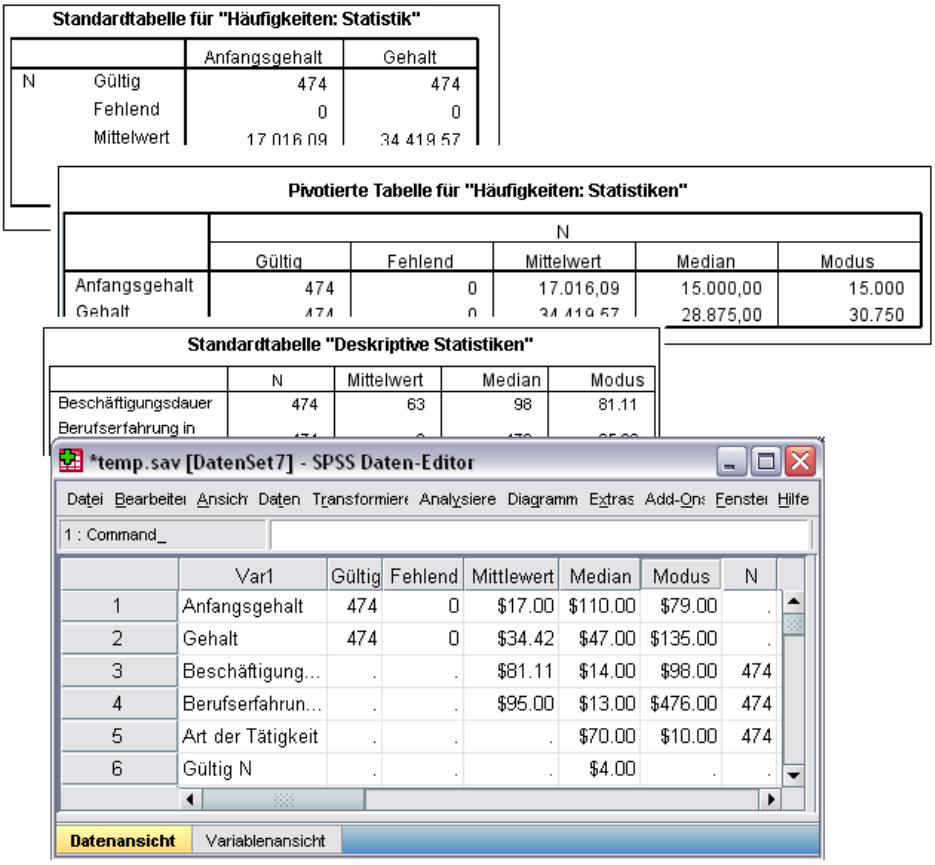


Abbildung 48-13
Kombinieren verschiedener Tabellentypen in einer Datendatei durch Pivotieren von Dimensionselementen



Standardtabelle für "Häufigkeiten: Statistik"

		Anfangsgehalt	Gehalt
N	Gültig	474	474
	Fehlend	0	0
	Mittelwert	17.016,09	34.419,57

Pivotierte Tabelle für "Häufigkeiten: Statistiken"

	N				
	Gültig	Fehlend	Mittelwert	Median	Modus
Anfangsgehalt	474	0	17.016,09	15.000,00	15.000
Gehalt	474	0	34.419,57	28.875,00	30.750

Standardtabelle "Deskriptive Statistiken"

	N	Mittelwert	Median	Modus
Beschäftigungsdauer	474	63	98	81.11
Berufserfahrung in				

*temp.sav [DatenSet7] - SPSS Daten-Editor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Diagramm Extras Add-Ons Fenster Hilfe

1 : Command_

	Var1	Gültig	Fehlend	Mittlewert	Median	Modus	N
1	Anfangsgehalt	474	0	$17.00	$110.00	$79.00	.
2	Gehalt	474	0	$34.42	$47.00	$135.00	.
3	Beschäftigung...	.	.	$81.11	$14.00	$98.00	474
4	Berufserfahrun...	.	.	$95.00	$13.00	$476.00	474
5	Art der Tätigkeit	.	.	.	$70.00	$10.00	474
6	Gültig N	.	.	.	$4.00	.	.

Datenansicht | Variablenansicht

Bei einigen Variablen sind fehlende Werte vorhanden, weil die Tabellenstrukturen immer noch nicht völlig identisch mit den Statistiken in den Spalten sind.

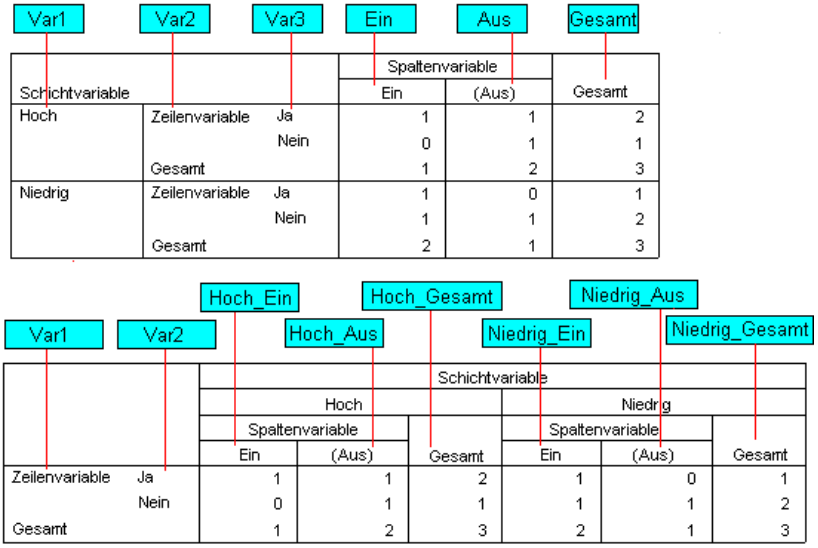
Variablennamen in Datendateien aus dem OMS

Das OMS erzeugt gültige, eindeutige Variablennamen aus den Spaltenbeschriftungen:

- Den Zeilen- und Schichtenelementen werden generische Variablennamen zugewiesen. Diese bestehen aus dem Präfix *Var* und einer laufenden Nummer.
- Zeichen, die in Variablennamen nicht zulässig sind (Leerzeichen, Klammern usw.) werden entfernt. “Diese (Spalten-)Beschriftung” würde beispielsweise zu einer Variablen mit der Bezeichnung *DieseSpalten-Beschriftung*.
- Ist das erste Zeichen der Beschriftung zwar grundsätzlich in Variablennamen zulässig, nicht jedoch als erstes Zeichen (z. B. Ziffern), wird das Symbol “@” als Präfix vorangestellt. “2tes” würde beispielsweise zu einer Variablen mit der Bezeichnung *@2tes*.

- Unterstriche und Punkte am Ende von Beschriftungen werden aus den resultierenden Variablennamen entfernt. (Die Unterstriche am Ende der automatisch erzeugten Variablen *Command_*, *Subtype_* und *Label_* bleiben erhalten.)
- Enthält die Spaltendimension mehrere Elemente, werden die Variablennamen aus einer Kombination der Kategorienbeschriftungen gebildet; die Kategorienbeschriftungen werden dabei durch einen Unterstrich getrennt. Gruppenbeschriftungen werden nicht berücksichtigt. Wenn beispielsweise *VarB* in *VarA* in den Spalten verschachtelt ist, erhalten Sie Variablen wie *CatA1_CatB1*, nicht jedoch *VarA_CatA1_VarB_CatB1*.

Abbildung 48-14
Variablennamen, die aus Tabellenelementen gebildet wurden



OXML-Tabellenstruktur

Output XML (OXML) ist XML, das dem Schema *spss-output* entspricht. Eine ausführliche Beschreibung des Schemas finden Sie im Abschnitt “Ausgabeschema” des Hilfesystems.

- Die Befehls- und Untertyp-IDs im OMS dienen als Werte für die Attribute command und subType in OXML. Ein Beispiel lautet folgendermaßen:

```
<command text=Häufigkeiten command=Häufigkeiten...>
 <pivotTable text=Geschlecht label=Geschlecht subType=Häufigkeiten...>
```

- Die Ausgabesprache sowie die Einstellungen für die Anzeige von Variablennamen/Beschriftungen und Werte/Wertelabels wirken sich nicht auf die Werte der OMS-Attribute command und subType aus.
- Bei XML wird zwischen Groß- und Kleinschreibung unterschieden. Der subType-Attributwert “häufigkeiten” ist *nicht* identisch mit dem subType-Attributwert “Häufigkeiten”.

- Alle in einer Tabelle angezeigten Informationen befinden sich in Attributwerten in OXML. Auf der Ebene einzelner Zellen besteht OXML aus “leeren” Elementen, die zwar Attribute enthalten, jedoch keine “Inhalte” (außer den Inhalten in den Attributwerten).
- Die Tabellenstruktur in OMXL wird zeilenweise dargestellt. Die Spaltenelemente sind in den Zeilen verschachtelt, und einzelne Zellen sind wiederum in den Spaltenelementen verschachtelt.

```
<pivotTable...>
 <dimension axis='row'...>
  <dimension axis='column'...>
   <category...>
    <cell text='...' number='...' decimals='...'/>
   </category>
   <category...>
    <cell text='...' number='...' decimals='...'/>
   </category>
  </dimension>
 </dimension>
  ...
 </pivotTable>
```

Das vorausgehende Beispiel ist eine vereinfachte Darstellung der Struktur, die die Beziehungen zwischen Nachfolgern und Vorgängern veranschaulicht. Das Beispiel zeigt jedoch nicht notwendigerweise die direkt über- oder untergeordneten Elemente, weil in der Regel verschachtelte Ebenen von Elementen vorliegen.

Die nachstehende Abbildung zeigt eine einfache Häufigkeitstabelle und die vollständige XML-Ausgabedarstellung dieser Tabelle.

Abbildung 48-15
Einfache Häufigkeitstabelle

Geschlecht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Männlich	258	54,4	54,4	54,4
	Weiblich	216	45,6	45,6	100,0
	Gesamt	474	100,0	100,0	

Abbildung 48-16
XML-Ausgabe für eine einfache Häufigkeitstabelle

```
<?xml version=1.0 encoding=UTF-8 ?>
<outputTreeoutputTree xmlns=http://xml.spss.com/spss/oms
 xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance
 xsi:schemaLocation=http://xml.spss.com/spss/oms
 http://xml.spss.com/spss/oms/spss-output-1.0.xsd>
<command text=Häufigkeiten command=Häufigkeiten
 displayTableValues=label displayOutlineValues=label
 displayTableVariables=label displayOutlineVariables=label>
 <pivotTable text=Geschlecht label=Geschlecht subType=Häufigkeiten
  varName=geschl variable=true>
```

```
<dimension axis=row text=Geschlecht label=Geschlecht
 varName=geschl variable=true>
 <group text=Gültig>
  <group hide=true text=Dummy>
   <category text=Weiblich label=Weiblich string=w
    varName=geschl>
    <dimension axis=column text=Statistik>
     <category text=Häufigkeit>
      <cell text=216 number=216/>
     </category>
     <category text=Prozent>
      <cell text=45,6 number=45,569620253165 decimals=1/>
     </category>
     <category text=Gültige Prozente>
      <cell text=45,6 number=45,569620253165 decimals=1/>
     </category>
     <category text=Kumulierte Prozente>
      <cell text=45,6 number=45,569620253165 decimals=1/>
     </category>
    </dimension>
   </category>
   <category text=Männlich label=Männlich string=m varName=geschl>
    <dimension axis=column text=Statistik>
     <category text=Häufigkeit>
      <cell text=258 number=258/>
     </category>
     <category text=Prozent>
      <cell text=54,4 number=54,430379746835 decimals=1/>
     </category>
     <category text=Gültige Prozente>
      <cell text=54,4 number=54,430379746835 decimals=1/>
     </category>
     <category text=Kumulierte Prozente>
      <cell text=100,0 number=100 decimals=1/>
     </category>
    </dimension>
   </category>
  </group>
  <category text=Gesamt>
   <dimension axis=column text=Statistik>
    <category text=Häufigkeit>
     <cell text=474 number=474/>
    </category>
    <category text=Prozent>
     <cell text=100,0 number=100 decimals=1/>
    </category>
    <category text=Gültige Prozente>
     <cell text=100,0 number=100 decimals=1/>
    </category>
   </dimension>
  </category>
 </group>
```

```
  </dimension>
 </pivotTable>
</command>
</outputTree>
```

Eine einfache, kleine Tabelle kann zu beträchtlichen Mengen an XML führen. Der Grund hierfür liegt teilweise darin, dass die XML einige Informationen enthält, die aus der ursprünglichen Tabelle nicht ohne weiteres ersichtlich sind, sowie einige Informationen, die in der ursprünglichen Tabelle nicht einmal vorlagen. Auch eine gewisse Redundanz ist vorhanden.

- Der Tabelleninhalt, wie er im Viewer in einer Pivot-Tabelle dargestellt wird (oder würde), ist in Textattributen enthalten. Ein Beispiel lautet folgendermaßen:

```
<command text=Häufigkeiten command=Häufigkeiten...>
```

- Die Ausgabesprache sowie die Einstellungen für die Anzeige von Variablennamen/Beschriftungen und Werte/Wertelabels wirken sich ggf. auf die Textattribute aus. In diesem Beispiel ist der Wert des Attributs text abhängig von der Ausgabesprache, der Wert des Attributs "command" bleibt dagegen unabhängig von der Ausgabesprache immer gleich.
- An allen Stellen, an denen Variablen oder Werte für Variablen in Zeilen- oder Spaltenbeschriftungen auftreten, enthält die XML ein Attribut text sowie mindestens einen weiteren Attributwert. Ein Beispiel lautet folgendermaßen:

```
<dimension axis=row text=Geschlecht label=Geschlecht varName=geschlecht>
 ...<category text=Weiblich label=Weiblich string=w varName=geschlecht>
```

- Bei einer numerischen Variable würde entsprechend ein Attribut number anstelle eines Attributs string verwendet. Das Attribut label ist nur dann vorhanden, wenn eine Beschriftung für die Variable oder die Werte definiert wurde.
- Die Elemente <cell> mit den Zellwerten für Zahlen enthalten das Attribut text sowie mindestens einen weiteren Attributwert. Ein Beispiel lautet folgendermaßen:

```
<cell text=45,6 number=45,569620253165 decimals=1/>
```

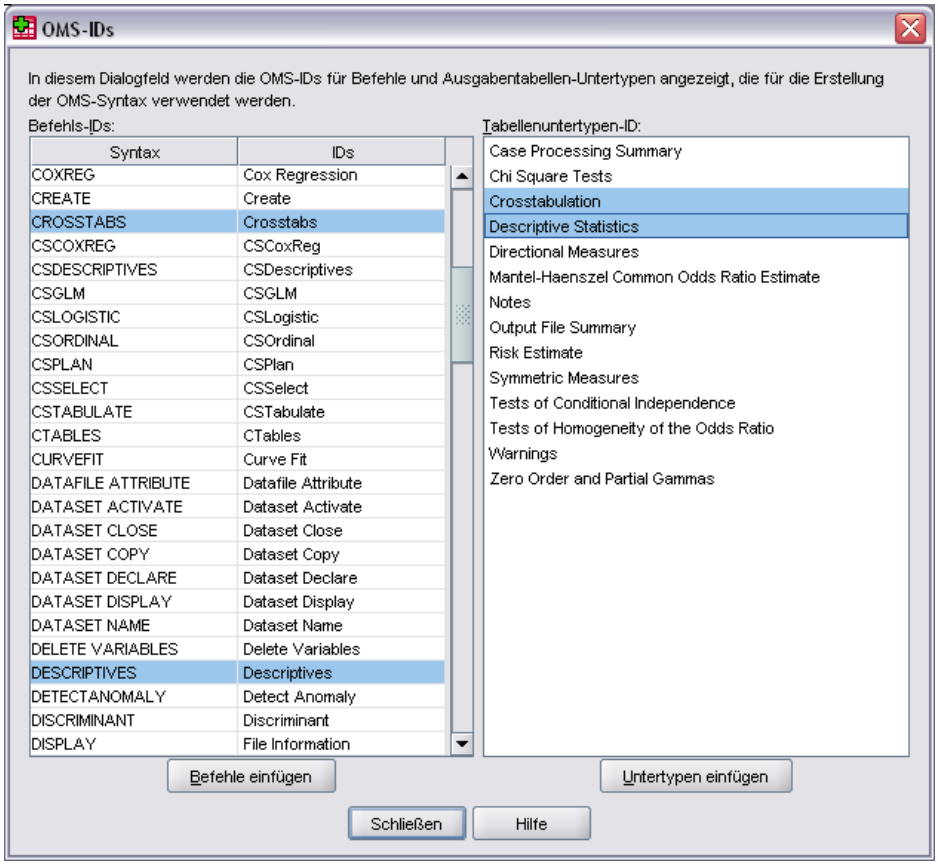
Das Attribut number ist der eigentliche, nicht gerundete numerische Wert, und das Attribut decimals bezeichnet die Anzahl der Dezimalstellen, die in der Tabelle angezeigt werden.

- Da die Spalten in den Zeilen verschachtelt sind, wird das Kategorieelement für die einzelnen Spalten für jede Zeile wiederholt. Beispiel: Die Statistik wird in den Spalten angezeigt, und daher wird das Element <category text=Häufigkeit> dreimal in der XML aufgeführt: einmal für die Zeile "Männlich", einmal für die Zeile "Weiblich" und einmal für die Zeile "Gesamt".

OMS-IDs

Das Dialogfeld “OMS-IDs” soll Sie beim Schreiben von OMS-Befehlssyntax unterstützen. Mit diesem Dialogfeld können Sie ausgewählte Befehls- und Untertypen-IDs in ein Befehlssyntax-Fenster einfügen.

Abbildung 48-17
Dialogfeld “OMS-IDs”



So verwenden Sie das Dialogfeld “OMS-IDs”:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Extras
  OMS-IDs...

- Wählen Sie einen oder mehrere Befehls- oder Untertypen-IDs aus. (Halten Sie beim Klicken mit der Maus gleichzeitig die Strg-Taste gedrückt, um mehrere IDs in den einzelnen Listen auszuwählen.)

- Klicken Sie auf Befehle einfügen und/oder Untertypen einfügen.
  - Die Liste der verfügbaren Untertypen hängt von den zum jeweiligen Zeitpunkt ausgewählten Befehlen ab. Wenn mehrere Befehle ausgewählt wurden, stellt die Liste der verfügbaren Untertypen die Gesamtheit aller Untertypen dar, die für die jeweils ausgewählten Befehle verfügbar sind. Wenn keine Befehle ausgewählt wurden, werden alle Untertypen aufgeführt.

- Die IDs werden in das Hauptfenster für die Befehlssyntax an der jeweiligen Cursorposition eingefügt. Wenn keine Befehlssyntax-Fenster geöffnet sind, wird automatisch ein neues Syntax-Fenster geöffnet.
- Jede Befehls- und/oder Untertypen-ID wird beim Einfügen in Anführungszeichen eingeschlossen, da diese Anführungszeichen für die Befehlssyntax von OMS erforderlich sind.
- ID-Listen für die Schlüsselwörter COMMANDS und SUBTYPES müssen in Klammern eingeschlossen sein, wie in folgendem Beispiel:

```
/IF COMMANDS=['Kreuztabellen' 'Deskriptive Statistiken']
    SUBTYPES=['Kreuztabelle' 'Deskriptive Statistiken']
```

Kopieren von OMS-IDs aus Viewer-Gliederung.

Sie können OMS-Befehls-IDs und OMS-Untertypen-IDs aus dem Viewer-Gliederungsfenster kopieren und einfügen.

▶ Klicken Sie im Gliederungsfenster mit der rechten Maustaste auf den Gliederungseintrag für das Element.

▶ Wählen Sie die Option OMS-Befehls-ID kopieren oder die Option OMS-Tabellenuntertyp kopieren.

Diese Methode unterscheidet sich in einem Punkt von der Verwendung des Dialogfelds "OMS-IDs": Die kopierte ID wird nicht automatisch in das Befehlssyntax-Fenster eingefügt. Die ID wird einfach in die Zwischenablage kopiert, und Sie können sie anschließend an jeder gewünschten Stelle einfügen. Da die Werte für die Befehls- und die Untertypen-IDs exakt mit den zugehörigen Befehls- und Untertypen-Attributwerten im Format "Ausgabe-XML" (OXML) übereinstimmen, ist diese Methode des Kopierens und Einfügens besonders hilfreich beim Schreiben von XSLT-Transformationen.

Kopieren von OMS-Labels

Statt der IDs können Sie Labels für die Verwendung mit dem Schlüsselwort LABELS kopieren. Labels können verwendet werden, um zwischen mehreren Diagrammen oder mehreren Tabellen desselben Typs zu unterscheiden, bei denen der Gliederungstext ein Attribut des jeweiligen Ausgabeobjekts angibt, beispielsweise die Variablennamen oder -labels. Es gibt jedoch eine Reihe von Faktoren, die den Labeltext beeinflussen können:

- Wenn die Verarbeitung aufgeteilter Dateien aktiviert ist, kann die Gruppen-ID für die aufgeteilte Datei an die Beschriftung angehängt werden.
- Labels, die Informationen über Variablen oder Werte enthalten, hängen unter anderem von den Einstellungen für die Anzeige von Variablennamen/-werten und den Werten/Wertelabels im Gliederungsfenster ab (Menü "Bearbeiten", "Optionen", Registerkarte "Beschriftung der Ausgabe").
- Labels richten sich außerdem nach der aktuellen Einstellung für die Ausgabesprache (Menü "Bearbeiten", "Optionen", Registerkarte "Allgemein").

So kopieren Sie OMS-Beschriftungen:

- Klicken Sie im Gliederungsfenster mit der rechten Maustaste auf den Gliederungseintrag für das Element.
- Wählen Sie OMS-Label kopieren aus.

  Wie bei den Befehls- und Untertypen-IDs müssen die Beschriftungen in Anführungszeichen und die gesamte Liste in eckige Klammern eingeschlossen sein, wie in folgendem Beispiel:

```
/IF LABELS=['Art der Tätigkeit' 'Schulabschluss']
```

Index

Abbrechen (Schaltfläche), 8
Abhängiger *T*-Test
  in T-Test bei gepaarten Stichproben, 316
Abweichungskontraste
  in GLM, 332–333
Access (Microsoft), 20
Aggregieren von Daten, 195
  Aggregierungsfunktionen, 198
  Variablennamen und -labels, 198
Ähnlichkeiten
  in der hierarchischen Clusteranalyse, 407
Ähnlichkeitsmaße
  in der hierarchischen Clusteranalyse, 409
  in Distanzen, 352
Aktive Datei, 68–69
  Erstellen einer temporären aktiven Datei, 69
  in Zwischenspeicher ablegen, 69
  virtuelle aktive Datei, 68
Aktives Fenster, 6
Algorithmen, 12
Alpha-Faktorisierung, 392
Alpha-Koeffizient
  in der Reliabilitätsanalyse, 456, 458
Analyse von Mehrfachantworten
  Häufigkeitstabellen, 438
  Kreuztabelle, 440
  Mehrfachantworten: Häufigkeiten, 438
  Mehrfachantworten: Kreuztabellen, 440
Anderson-Rubin-Faktorwerte, 395
Andrew-Wellen-Schätzer
  in der Explorativen Datenanalyse, 286
Anfänglicher Schwellenwert
  in der Two-Step-Clusteranalyse, 402
Anmelden bei einem Server, 72
ANOVA
  in einfaktorieller ANOVA, 321
  in GLM - Univariat, 328
  in "Mittelwerte", 304
  Modell, 330
Anteilsschätzungen
  beim Bilden der Rangfolge, 153
Anzahl der Fälle
  in "Mittelwerte", 304
  in OLAP-Würfel, 309
  in Zusammenfassen, 299
Anzeigeformate, 86
Anzeigereihenfolge, 246
Arbeitsspeicher, 491
Attribute
  Benutzerdefinierte Variablenattribute, 91
Auflisten von Fällen, 297
Aufteilen von Tabellen, 267
  Festlegen von Tabellenumbrüchen, 267
Ausblenden, 225, 251–252, 264, 509
  Dimensionsbeschriftungen, 252
  Erklärungen, 264
  Fußnoten, 264
  Prozedurergebnisse, 225
  Symbolleisten, 509
  Titel, 252
  Zeilen und Spalten, 251
Äußere Verbindung, 24
Ausgabe, 224–226, 231–232, 244, 493
  Ändern der Ausgabesprache, 491
  Ausblenden, 225
  Ausrichtung, 226, 493
  Einblenden, 225
  Einfügen in andere Anwendungen, 231
  exportieren, 232
  Kopieren, 225
  kopieren in andere Anwendungen, 231
  löschen, 225–226
  speichern, 244
  Verschieben, 225
  Viewer, 224
  zentrieren, 226, 493
Ausgabeobjekttypen
  bei OMS, 526
Ausgabeverwaltungssystem (OMS), 523, 547
Ausgeschlossene Residuen
  in GLM, 337
  in Lineare Regression, 359
Ausreißer
  in der Explorativen Datenanalyse, 286
  in der Two-Step-Clusteranalyse, 402
  in Lineare Regression, 357
Ausrichtung, 89, 226, 493
  Ausgabe, 226, 493
  im Daten-Editor, 89
Ausschließen der Ausgabe aus dem Viewer bei OMS, 534
Auswählen von Fällen, 200
  auf der Grundlage von Auswahlkriterien, 202
  Bereich von Fällen, 204
  Datumsbereich, 204
  Zeitbereich, 204
  Zufallsstichprobe, 203
Auswahlmethoden, 266
  Auswählen von Zeilen und Spalten in Pivot-Tabellen, 266
Auswahlvariable
  in Lineare Regression, 357
automatisierte Produktion, 514
Autoskripts, 505

Balkendiagramme
in Häufigkeiten, 278
Bartlett-Faktorwerte, 395
Bartlett-Test auf Sphärizität
in der Faktorenanalyse, 391
Baumtiefe
in der Two-Step-Clusteranalyse, 402
Bearbeiten von Daten, 97–98
Bedingte Transformationen, 142
Befehls-IDs, 527
Befehlssprache, 269
Befehlssyntax, 269, 273, 508, 512, 514
ausführen, 273
Ausführen mit Symbolleisten-Schaltflächen, 512
Ausgabe-Log, 271
einfügen, 271
Hinzufügen zu Menüs, 508
Journaldatei, 274
Produktionsjobs, Regeln, 514
Syntaxregeln, 269
Zugreifen auf *SPSS Command Syntax Reference*, 12
Befehlssyntaxdateien, 273
Befehlszeilenschalter, 521
Produktionsjobs, 521
Benutzerdefinierte Attribute, 91
Benutzerdefinierte fehlende Werte, 88
Benutzerdefinierte Modelle
in GLM, 330
Benutzerdefinierte Variablenattribute, 91
Beobachtete Anzahl
in Kreuztabellen, 295
Beobachtete Häufigkeiten
in Ordinale Regression, 368
Beobachtete Mittelwerte
in GLM - Univariat, 338
Berechnen von Variablen, 140
Berechnen von neuen String-Variablen, 142
Bereich
in Deskriptive Statistiken, 281
in Häufigkeiten, 276
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Verhältnisstatistiken, 469
in Zusammenfassen, 299
Bereichseinteiler, 131
Bereichseinteilung, 131
Bericht in Spalten, 450
Fehlende Werte, 454
Gesamtergebnis, 454
Gesamtergebnisspalten, 453
Seiteneinstellung, 454
Seitenformat, 449
Seitennumerierung, 454
Spaltenformat, 446
zusätzliche Funktionen beim Befehl, 455
Zwischenergebnisse, 454
Bericht in Zeilen, 444
Break-Abstand, 447
Break-Spalten, 444
Datenspalten, 444
Fehlende Werte, 448
Fußzeilen, 450
Seiteneinstellung, 447
Seitenformat, 449
Seitennumerierung, 448
Sortierfolgen, 444
Spaltenformat, 446
Titel, 450
Variablen in Titel, 450
zusätzliche Funktionen beim Befehl, 455
Berichte
Berichte in Spalten, 450
Berichte in Zeilen, 444
Dividieren von Spaltenwerten, 453
Gesamtergebnisspalten, 453
Multiplizieren von Spaltenwerten, 453
Vergleichen von Spalten, 453
zusammengesetzte Gesamtergebnisse, 453
Berichte in Spalten, 450
Beschriftungen, 247
Einfügen von Gruppenbeschriftungen, 247
im Vergleich mit Untertypennamen in OMS, 528
löschen, 247
Beta-Koeffizienten
in Lineare Regression, 362
Bewertung
Anzeigen der geladenen Modelle, 183
Für Export und Bewertung unterstützte Modelle, 180
Laden von gespeicherten Modellen, 181
Bilden der Rangfolge, 152
gebundene Werte, 154
Perzentile, 153
relative Ränge, 153
Savage-Werte, 153
Bivariate Korrelationen
Fehlende Werte, 343
Korrelationskoeffizienten, 341
Optionen, 343
Signifikanzniveau, 341
Statistiken, 343
zusätzliche Funktionen beim Befehl, 343
Block-Distanz
in Distanzen, 351
Blom-Schätzungen, 153
BMP-Dateien, 232, 239
Exportieren von Diagrammen, 232, 239
Bonferroni
in einfaktorieller ANOVA, 323
in GLM, 335
Box-M-Test
in der Diskriminanzanalyse, 384
Boxplots
in der Explorativen Datenanalyse, 287

Vergleichen von Faktorstufen, 287
Vergleichen von Variablen, 287
Break-Variablen
in Aggregieren von Daten, 195
Breite Tabellen
Einfügen in Microsoft Word, 231
Brown-Forsythe-Statistik
in einfaktorieller ANOVA, 326

C nach Dunnett
in einfaktorieller ANOVA, 323
in GLM, 335
chartsoutput, 231
Chi-Quadrat, 417
auf Unabhängigkeit, 292
erwartete Werte, 419
erwarteter Bereich, 419
Exakter Test nach Fisher, 292
Fehlende Werte, 419
in Kreuztabellen, 292
Kontinuitätskorrektur nach Yates, 292
Likelihood-Quotient, 292
Optionen, 419
Pearson-Korrelationskoeffizient, 292
Statistiken, 419
Test bei einer Stichprobe, 417
Zusammenhang linear-mit-linear, 292
Chi-Quadrat-Distanz
in Distanzen, 351
Cluster-Häufigkeiten
in der Two-Step-Clusteranalyse, 405
Clusteranalyse
Auswählen einer Prozedur, 397
Clusterzentrenanalyse, 412
Effizienz, 414
Hierarchische Clusteranalyse, 407
Clusterzentrenanalyse
Beispiele, 412
Cluster-Zugehörigkeit, 415
Distanzen der Cluster, 415
Effizienz, 414
Fehlende Werte, 415
Iterationen, 414
Konvergenzkriterien, 414
Methoden, 412
Speichern von Cluster-Informationen, 415
Statistiken, 412, 415
Übersicht, 412
zusätzliche Funktionen beim Befehl, 416
Cochran-*Q*
in Tests bei mehreren verbundenen Stichproben, 435
Cochran-Statistik
in Kreuztabellen, 292
Cohen-Kappa
in Kreuztabellen, 292
Cook-Distanz
in GLM, 337
in Lineare Regression, 359
Cox/Snell-R^2
in Ordinale Regression, 368
Cramér-*V*
in Kreuztabellen, 292
Cronbachs Alpha
in der Reliabilitätsanalyse, 456, 458
CSV-Format
Einlesen von Daten, 33
Speichern von Daten, 47

d
in Kreuztabellen, 292
DATA LIST, 68
Vergleich mit GET DATA, 68
Datei-Speicherstellen
Festlegen der standardmäßigen Datei-Speicherstellen, 504
Dateien, 229
Hinzufügen von Textdateien im Viewer, 229
öffnen, 15
Dateitransformationen, 206
Aggregieren von Daten, 195
Gewichten von Fällen, 204
Sortieren von Fällen, 185
Transponieren von Variablen und Fällen, 188
Umstrukturieren von Daten, 206
Verarbeitung von aufgeteilten Dateien, 199
Zusammenfügen von Datendateien, 189, 192
Daten aus Dimensions, 42
speichern, 66
Daten exportieren, 47
Daten importieren, 15, 19
Daten-Editor, 80, 82, 89, 96–100, 102–103, 508
Ändern des Datentyps, 100
Anzeigeoptionen, 103
Ausrichtung, 89
Bearbeiten von Daten, 97–98
Datenansicht, 80
Definieren von Variablen, 82
Drucken, 103
Einfügen von neuen Fällen, 99
Einfügen von neuen Variablen, 99
Eingeben von Daten, 96
Eingeben von nichtnumerischen Daten, 97
Eingeben von numerischen Daten, 96
Einschränkungen für die Datenwerte, 97
Gefilterte Fälle, 102
mehrere Ansichten/Fenster, 103
Mehrere geöffnete Datendateien, 105, 491
Senden von Daten an andere Anwendungen, 508
Spaltenbreite, 89
Variablenansicht, 81
Verschieben von Variablen, 100
Daten-Sets
umbenennen, 108

Datenanalyse, 10
  grundlegende Schritte, 10
Datenansicht, 80
Datenbanken, 19–20, 22, 24–25, 28, 30, 32
  Aktualisieren, 54
  Anhängen von Datensätzen (Fällen) zu einer Tabelle, 63
  Auswählen einer Datenquelle, 20
  Auswählen von Datenfeldern, 22
  bedingte Ausdrücke, 25
  Definieren von Variablen, 30
  einlesen, 19–20, 22
  Ersetzen einer Tabelle, 64
  Ersetzen von Werten in bestehenden Feldern, 61
  Erstellen einer neuen Tabelle, 64
  Erstellen von Beziehungen, 24
  Festlegen von Kriterien, 25
  Hinzufügen neuer Felder zu einer Tabelle, 62
  Microsoft Access, 20
  Parameterabfragen, 25, 28
  speichern, 54
  Speichern von Abfragen, 32
  SQL-Syntax, 32
  Überprüfen von Ergebnissen, 32
  Umwandeln von Strings in numerische Variablen, 30
  Verbindungen zwischen Tabellen, 24
  Wert abfragen, 28
  Where-Klausel, 25
  Zufallsstichproben, 25
Datendateien, 15–16, 33, 46–47, 53, 69, 76, 206
  Dimensions, 42
  Hinzufügen von Kommentaren, 486
  Informationen aus dem Datenlexikon, 46
  Informationen zur Datei, 46
  Leistungssteigerung bei umfangreichen Dateien, 69
  Mehrere geöffnete Datendateien, 105, 491
  mrInterview, 42
  öffnen, 15–16
  Quancept, 42
  Quanvert, 42
  Remote-Server, 76
  schützen, 67
  speichern, 46–47
  Speichern von Ausgaben als Datendateien im SPSS-Format, 523
  Speichern von Untergruppen von Variablen, 53
  Text, 33
  transponieren, 188
  umstrukturieren, 206
  vertauschen, 188
Dateneingabe, 96
Datenlexikon, 46
  Zuweisen aus einer anderen Datei, 120
Datentransformationen, 494
  Bedingte Transformationen, 142
  Berechnen von Variablen, 140
  Bilden der Rangfolge, 152
  Funktionen, 143
  String-Variablen, 142
  Umkodieren von Werten, 147–148, 150–151, 155
  verzögerte Ausführung, 494
  Zeitreihen, 173, 175
Datentypen, 84, 86, 100, 496
  ändern, 100
  Anzeigeformate, 86
  definieren, 84
  Eingabeformate, 86
  spezielle Währung, 84, 496
Datumsformate
  zweistellige Jahresangaben, 494
Datumsformatvariablen, 84, 86, 494
  Addieren oder Subtrahieren zu bzw. von Datums-/Zeitvariablen, 158
  Erstellen einer Datums-/Zeitvariablen aus einem String, 158
  Erstellen einer Datums-/Zeitvariablen aus einem Variablen-Set, 158
  Extrahieren eines Teils einer Datums-/Zeitvariablen, 158
Datumsvariablen
  für Zeitreihendaten definieren, 174
dBASE-Dateien, 15, 18, 47
  einlesen, 15, 18
  speichern, 47
Definieren von Variablen, 82, 84, 87–90, 111
  Datentypen, 84
  Fehlende Werte, 88
  Kopieren und Einfügen von Attributen, 89–90
  Variablenlabels, 87
  Vorlagen, 89–90
  Wertelabels, 87, 111
  Zuweisen eines Datenlexikons, 120
Dendrogramme
  in der hierarchischen Clusteranalyse, 410
Deskriptive Statistiken, 280
  Anzeigereihenfolge, 281
  in der Explorativen Datenanalyse, 286
  in der Two-Step-Clusteranalyse, 405
  in Deskriptive Statistiken, 280
  in GLM - Univariat, 338
  in Häufigkeiten, 276
  in Verhältnisstatistiken, 469
  in Zusammenfassen, 299
  Speichern von *Z*-Werten, 280
  Statistiken, 281
  zusätzliche Funktionen beim Befehl, 283
DfBeta
  in Lineare Regression, 359
DfFit
  in Lineare Regression, 359
Diagramm-Editor, 478
  Eigenschaften, 479
Diagramme, 225, 232, 268, 474, 499
  Ausblenden, 225
  Diagrammerstellung, 474
  erstellen, 474

Erstellen aus Pivot-Tabellen, 268
exportieren, 232
Fallbeschriftungen, 372
Fehlende Werte, 482
Größe, 482
in ROC-Kurve, 471
kopieren in andere Anwendungen, 231
Seitenverhältnis, 499
Übersicht, 474
Umbrechen von Feldern, 482
Vorlagen, 482, 499
Diagramme mit der Streubreite gegen das mittlere Niveau
in der Explorativen Datenanalyse, 287
in GLM - Univariat, 338
Diagrammerstellung, 474
Galerie, 475
Diagrammoptionen, 499
Dialogfelder, 9, 486–487, 491
Anzeigen von Variablenlabels, 7, 491
Anzeigen von Variablennamen, 7, 491
Anzeigereihenfolge für Variablen, 491
Auswählen von Variablen, 9
Definieren von Variablen-Sets, 486
Steuerelemente, 8
Umsortieren von Listen der Zielvariablen, 489
Variablen, 7
Variablenbeschreibung, 10
Variablensymbole, 9
Verwenden von Variablen-Sets, 487
Differenzen zwischen Gruppen
in OLAP-Würfel, 311
Differenzen zwischen Variablen
in OLAP-Würfel, 311
Differenzfunktion, 176
Differenzkontraste
in GLM, 332–333
Direkte Oblimin-Rotation
in der Faktorenanalyse, 394
Diskriminanzanalyse, 381
A-priori-Wahrscheinlichkeit, 386
Anzeigeoptionen, 385–386
Auswählen von Fällen, 383
Beispiel, 381
Definieren eines Bereichs, 383
Deskriptive Statistiken, 384
Diagramme, 386
Diskriminanzmethoden, 385
Exportieren von Modellinformationen, 388
Fehlende Werte, 386
Funktionskoeffizienten, 384
Gruppenvariablen, 381
Kovarianzmatrix, 386
Kriterien, 385
Mahalanobis-Abstand, 385
Matrizen, 384
Rao-*V*, 385
schrittweise Methoden, 381
Speichern von Klassifikationsvariablen, 388
Statistiken, 381, 384
unabhängige Variablen, 381
Wilks-Lambda, 385
zusätzliche Funktionen beim Befehl, 388
Distanz nach Minkowski
in Distanzen, 351
Distanz nach Tschebyscheff
in Distanzen, 351
Distanzen, 349
Ähnlichkeitsmaße, 352
Beispiel, 349
Berechnen von Distanzen zwischen Fällen, 349
Berechnen von Distanzen zwischen Variablen, 349
Statistiken, 349
Transformieren von Maßen, 351–352
Transformieren von Werten, 351–352
Unähnlichkeitsmaße, 351
zusätzliche Funktionen beim Befehl, 353
Distanzmaße
in der hierarchischen Clusteranalyse, 409
in Distanzen, 351
Division
Dividieren über Berichtsspalten, 453
Dollarformat (DOLLAR), 84, 86
Doppelte Fälle (Datensätze)
finden und filtern, 127
Drehen von Beschriftungen, 248
Drucken, 103, 240–241, 243, 254, 259, 267
Abstand zwischen Ausgabeobjekten, 243
Daten, 103
Diagramme, 240
Diagrammgröße, 243
Festlegen von Tabellenumbrüchen, 267
Kopf- und Fußzeilen, 241
Pivot-Tabellen, 240
Schichten, 240, 254, 259
Seitenansicht, 240
Seitennummern, 243
Skalieren von Tabellen, 254, 259
Textausgabe, 240
Duncans multipler Spannweitentest
in einfaktorieller ANOVA, 323
in GLM, 335
Dunnett-*T*-Test
in einfaktorieller ANOVA, 323
in GLM, 335
Durbin-Watson-Statistik
in Lineare Regression, 362
Durchschnittliche absolute Abweichung (AAD)
in Verhältnisstatistiken, 469

Ehrlich signifikante Differenz nach Tukey
in einfaktorieller ANOVA, 323
in GLM, 335
Eigenschaften, 254
Pivot-Tabellen, 254

Tabellen, 254
Eigenwerte
in der Faktorenanalyse, 391–392
in Lineare Regression, 362
Einblenden, 225, 251–252, 264, 509
Dimensionsbeschriftungen, 252
Ergebnisse, 225
Erklärungen, 264
Fußnoten, 264
Symbolleisten, 509
Titel, 252
Zeilen oder Spalten, 251
Einfache Kontraste
in GLM, 332–333
Einfaktorielle ANOVA, 321
Faktorvariablen, 321
Fehlende Werte, 326
Kontraste, 322
Mehrfachvergleiche, 323
Optionen, 326
Polynomiale Kontraste, 322
Post-Hoc-Tests, 323
Statistiken, 326
zusätzliche Funktionen beim Befehl, 327
Einfügen (Schaltfläche), 8
Einfügen von Gruppenbeschriftungen, 247
Eingabeformate, 86
Eingeben von Daten, 96–97
nichtnumerisch, 97
Numerisch, 96
Verwenden von Wertelabels, 97
Eiszapfendiagramme
in der hierarchischen Clusteranalyse, 410
Entfernen von Gruppenbeschriftungen, 247
EPS-Dateien, 232, 239
Exportieren von Diagrammen, 232, 239
Equamax-Rotation
in der Faktorenanalyse, 394
Erklärungen, 263–264
Ersetzen fehlender Werte
lineare Interpolation, 179
linearer Trend, 179
Median der Nachbarpunkte, 179
Mittel der Nachbarpunkte, 179
Mittelwert der Datenreihe, 179
Erste
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Erwartete Anzahl
in Kreuztabellen, 295
Erwartete Häufigkeiten
in Ordinale Regression, 368
Eta
in Kreuztabellen, 292
in “Mittelwerte”, 304
Eta-Quadrat
in GLM - Univariat, 338
in “Mittelwerte”, 304
Euklidische Distanz
in Distanzen, 351
Exakter Test nach Fisher
in Kreuztabellen, 292
Excel-Dateien, 15, 17, 47, 508
Hinzufügen eines Menüeintrags zum Senden von Daten an Excel, 508
öffnen, 15, 17
speichern, 47
Speichern von Variablenlabels anstatt von Werten, 47
Excel-Format
Exportieren von Ausgaben, 232, 234
EXECUTE (Befehl)
Einfügen aus Dialogfeldern, 274
Explorative Datenanalyse, 284
Diagramme, 287
Fehlende Werte, 288
Optionen, 288
Potenztransformationen, 288
Statistiken, 286
zusätzliche Funktionen beim Befehl, 289
Exponentielles Modell
in Kurvenanpassung, 374
Exportieren von Ausgaben, 232, 235, 237
Excel-Format, 232, 234
HTML, 234
HTML-Format, 232
OMS, 523
PDF-Format, 232, 235
PowerPoint-Format, 232
Word-Format, 232, 234
Exportieren von Daten, 508
Hinzufügen eines Menüeintrags zum Exportieren von Daten, 508
Exportieren von Diagrammen, 232, 238–239, 514
automatisierte Produktion, 514
Extremwerte
in der Explorativen Datenanalyse, 286

F nach R-E-G-W
in einfaktorieller ANOVA, 323
in GLM, 335
Faktorenanalyse, 389
Anzeigeformat für Koeffizienten, 396
Auswählen von Fällen, 390
Beispiel, 389
deskriptive Statistiken, 391
Extraktionsmethoden, 392
Faktorwerte, 395
Fehlende Werte, 396
Konvergenz, 392, 394
Ladungsdiagramme, 394
Rotationsmethoden, 394
Statistiken, 389, 391

Übersicht, 389
zusätzliche Funktionen beim Befehl, 396
Faktorwerte, 395
Fälle, 99, 206
Auffinden doppelt vorhandener, 127
Auswählen von Teilmengen, 200, 202, 204
Einfügen von neuen Fällen, 99
Gewichtung, 204
sortieren, 185
Suchen im Daten-Editor, 100
Umstrukturieren in Variablen, 206
Fälle auswählen, 200
Fallweise Diagnose
in Lineare Regression, 362
Farben in Pivot-Tabellen, 258
Rahmen, 258
Fehlende Werte, 88, 482
definieren, 88
Diagramme, 482
im Sequenzentest, 423
in Bericht in Zeilen, 448
in Berichte in Spalten, 454
in bivariaten Korrelationen, 343
in Chi-Quadrat-Test, 419
in der Explorativen Datenanalyse, 288
in der Faktorenanalyse, 396
in einfaktorieller ANOVA, 326
in Funktionen, 143
in Kolmogorov-Smirnov-Test bei einer Stichprobe, 425
in Lineare Regression, 363
in Mehrfachantworten: Häufigkeiten, 438
in Mehrfachantworten: Kreuztabellen, 442
in Partielle Korrelationen, 347
in ROC-Kurve, 472
in T-Test bei einer Stichprobe, 319
in T-Test bei gepaarten Stichproben, 317
in T-Test bei unabhängigen Stichproben, 316
in Test auf Binomialverteilung, 421
in Tests bei mehreren unabhängigen Stichproben, 433
in Tests bei zwei unabhängigen Stichproben, 428
in Tests bei zwei verbundenen Stichproben, 431
in Zeitreihendaten ersetzen, 178
String-Variablen, 88
Fenster, 4
Aktives Fenster, 6
Hauptfenster, 6
Fensterteiler
Daten-Editor, 103
Festes Format, 33
Formatierung
Spalten in Berichten, 446
Fortsetzungstext, 259
für Pivot-Tabellen, 259
Freies Format, 33
Friedman-Test
in Tests bei mehreren verbundenen Stichproben, 435
Funktion für gleitenden Median, 176
Funktion für kumulierte Summe, 176
Funktion für saisonale Differenz, 176
Funktion für zentrierten gleitenden Durchschnitt, 176
Funktion für zurückgreifenden gleitenden Durchschnitt, 176
Funktionen, 143
Behandlung fehlender Werte, 143
Fußnoten, 255, 263–265
Diagramme, 481
Markierungen, 255, 264
neu numerieren, 265
Fußzeilen, 241

Gamma
in Kreuztabellen, 292
Gefilterte Fälle, 102
im Daten-Editor, 102
Geometrisches Mittel
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Geringste signifikante Differenz
in einfaktorieller ANOVA, 323
in GLM, 335
Gesamtergebnisse
in Berichte in Spalten, 454
Gesamtergebnisspalte
in Berichten, 453
Gesamtprozentwerte
in Kreuztabellen, 295
Gesättigte Modelle
in GLM, 330
Geschätzte Randmittel
in GLM - Univariat, 338
Geschwindigkeit, 69
Zwischenspeichern von Daten, 69
GET DATA, 68
Vergleich mit DATA LIST, 68
Vergleich mit GET CAPTURE, 68
Getrimmtes Mittel
in der Explorativen Datenanalyse, 286
Gewichten von Fällen, 204
nichtganzzahlige Gewichtungen in Kreuztabellen, 204
Gewichtete Daten, 222
und umstrukturierte Datendateien, 222
Gewichtete kleinste Quadrate
in Lineare Regression, 354
Gewichtete Schätzwerte
in GLM, 337
Gewichteter Mittelwert
in Verhältnisstatistiken, 469
Gitterlinien, 265
Pivot-Tabellen, 265
Glättungsfunktion, 176
Gliederung, 226–227
Ändern von Ebenen, 227
erweitern, 227

im Viewer, 226
reduzieren, 227
GLM
Modell, 330
Post-Hoc-Tests, 335
Profilplots, 334
Quadratsumme, 330
Speichern von Matrizen, 337
Speichern von Variablen, 337
GLM - Univariat, 328, 340
anzeigen, 338
Diagnose, 338
Geschätzte Randmittel, 338
Kontraste, 332–333
Optionen, 338
Goodman-und-Kruskal-Gamma
in Kreuztabellen, 292
Goodman-und-Kruskal-Lambda
in Kreuztabellen, 292
Goodman-und-Kruskal-Tau
in Kreuztabellen, 292
Größen, 228
in der Gliederung, 228
Größendifferenzmaß
in Distanzen, 351
grundlegende Schritte, 10
Gruppenbeschriftungen, 247
Gruppenmittelwerte, 302, 307
Gruppenvariablen, 206
erstellen, 206
Gruppieren von Zeilen oder Spalten, 247
Gruppierter Median
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
GT2 nach Hochberg
in einfaktorieller ANOVA, 323
in GLM, 335
Güte der Anpassung
in Ordinale Regression, 368
Guttman-Modelle
in der Reliabilitätsanalyse, 456, 458

Harmonisches Mittel
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Häufigkeiten, 275
Anzeigereihenfolge, 279
Diagramme, 278
Formate, 279
Statistiken, 276
Unterdrücken von Tabellen, 279
Häufigkeitstabellen
in der Explorativen Datenanalyse, 286
in Häufigkeiten, 275
Hauptachsen-Faktorenanalyse, 392
Hauptfenster, 6
Hauptkomponentenanalyse, 389, 392
Hebelwerte
in GLM, 337
in Lineare Regression, 359
Helmert-Kontraste
in GLM, 332–333
Hierarchische Clusteranalyse, 407
Ähnlichkeitsmaße, 409
Beispiel, 407
Cluster-Methoden, 409
Cluster-Zugehörigkeit, 410–411
Clustern von Fällen, 407
Clustern von Variablen, 407
Dendrogramme, 410
Diagrammausrichtung, 410
Distanzmaße, 409
Distanzmatrizen, 410
Eiszapfendiagramme, 410
Speichern von neuen Variablen, 411
Statistiken, 407, 410
Transformieren von Maßen, 409
Transformieren von Werten, 409
Zuordnungsübersichten, 410
zusätzliche Funktionen beim Befehl, 411
Hierarchische Zerlegung, 331
Hilfe (Schaltfläche), 8
Hilfe-Fenster, 12
Hintergrundfarbe, 261
Hinzufügen von Gruppenbeschriftungen, 247
Histogramme
in der Explorativen Datenanalyse, 287
in Häufigkeiten, 278
in Lineare Regression, 357
Höchstzahl Verzweigungen
in der Two-Step-Clusteranalyse, 402
Hotellings T^2
in der Reliabilitätsanalyse, 456, 458
HTML, 232, 234
Exportieren von Ausgaben, 232, 234

ICC. *Siehe* Korrelationskoeffizienten in Klassen, 458
Image-Faktorisierung, 392
in Zwischenspeicher ablegen, 69
Aktive Datei, 69
Informationen zur Datei, 46
Innere Verbindung, 24
Interaktive Diagramme, 231
kopieren in andere Anwendungen, 231
Inverses Modell
in Kurvenanpassung, 374
Iterationen
in der Clusterzentrenanalyse, 414
in der Faktorenanalyse, 392, 394
Iterationsprotokoll
in Ordinale Regression, 368

Jahre, 494
  zweistellige Angaben, 494
Journaldatei, 504
JPEG-Dateien, 232, 238
  Exportieren von Diagrammen, 232, 238

Kappa
  in Kreuztabellen, 292
Kategoriale Daten, 114
  Umwandeln von Intervalldaten in diskrete Kategorien, 131
Kendall-Tau-*b*
  in bivariaten Korrelationen, 341
  in Kreuztabellen, 292
Kendall-Tau-*c*, 292
  in Kreuztabellen, 292
Kendall-*W*
  in Tests bei mehreren verbundenen Stichproben, 435
Klassieren, 131
Klassifikation
  in ROC-Kurve, 471
Kollinearitätsdiagnose
  in Lineare Regression, 362
Kolmogorov-Smirnov-Test bei einer Stichprobe, 424
  Fehlende Werte, 425
  Optionen, 425
  Statistiken, 425
  zu testende Verteilung, 424
  zusätzliche Funktionen beim Befehl, 426
Kolmogorov-Smirnov-*Z*
  in Kolmogorov-Smirnov-Test bei einer Stichprobe, 424
  in Tests bei zwei unabhängigen Stichproben, 427
Kommaformat (COMMA), 84, 86
Kommagetrennte Dateien, 33
Konfidenzintervalle
  in der Explorativen Datenanalyse, 286
  in einfaktorieller ANOVA, 326
  in GLM, 332, 338
  in Lineare Regression, 362
  in ROC-Kurve, 472
  in T-Test bei einer Stichprobe, 319
  in T-Test bei gepaarten Stichproben, 317
  in T-Test bei unabhängigen Stichproben, 316
  Speichern in Lineare Regression, 359
Konstruieren von Termen, 331, 371
Kontingenzkoeffizient
  in Kreuztabellen, 292
Kontingenztafeln, 290
Kontinuitätskorrektur nach Yates
  in Kreuztabellen, 292
Kontraste
  in einfaktorieller ANOVA, 322
  in GLM, 332–333
Kontroll-Variablen
  in Kreuztabellen, 292
Konvergenz
  in der Clusterzentrenanalyse, 414
  in der Faktorenanalyse, 392, 394
Konzentrationsindex
  in Verhältnisstatistiken, 469
Kopfzeilen, 241
Korrelationen
  in bivariaten Korrelationen, 341
  in Kreuztabellen, 292
  in Partielle Korrelationen, 345
  nullter Ordnung, 347
Korrelationen nullter Ordnung
  in Partielle Korrelationen, 347
Korrelationskoeffizient nach Spearman
  in bivariaten Korrelationen, 341
  in Kreuztabellen, 292
Korrelationskoeffizienten in Klassen (ICC)
  in der Reliabilitätsanalyse, 458
Korrelationsmatrix
  in der Diskriminanzanalyse, 384
  in der Faktorenanalyse, 389, 391
  in Ordinale Regression, 368
korrigiertes R^2
  in Lineare Regression, 362
Kovarianzmatrix
  in der Diskriminanzanalyse, 384, 386
  in GLM, 337
  in Lineare Regression, 362
  in Ordinale Regression, 368
Kovarianzverhältnis
  in Lineare Regression, 359
KR20
  in der Reliabilitätsanalyse, 458
Kreisdiagramme
  in Häufigkeiten, 278
Kreuztabelle
  in Kreuztabellen, 290
  Mehrfachantworten, 440
Kreuztabellen, 290
  Formate, 296
  Gruppiertes Balkendiagramm, 292
  Kontroll-Variablen, 292
  nichtganzzahlige Gewichtungen, 204
  Schichten, 292
  Statistiken, 292
  Unterdrücken von Tabellen, 290
  Zellen anzeigen, 295
Kruskal-Tau
  in Kreuztabellen, 292
Kruskal-Wallis-*H*
  in Tests bei zwei unabhängigen Stichproben, 431
Kubisches Modell
  in Kurvenanpassung, 374
Kuder-Richardson-20 (KR20)
  in der Reliabilitätsanalyse, 458
Kumulative Häufigkeiten
  in Ordinale Regression, 368
Kurtosis
  in Bericht in Spalten, 452

in Bericht in Zeilen, 447
in der Explorativen Datenanalyse, 286
in Deskriptive Statistiken, 281
in Häufigkeiten, 276
in "Mittelwerte", 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Kurvenanpassung, 372
Einschließen von Konstanten, 372
Modelle, 374
Prognose, 375
Speichern von Residuen, 375
Speichern von Vorhersageintervallen, 375
Speichern vorhergesagter Werte, 375
Varianzanalyse, 372

Ladungsdiagramme
in der Faktorenanalyse, 394
LAG (Funktion), 176
Lagemaße
in der Explorativen Datenanalyse, 286
in Häufigkeiten, 276
in Verhältnisstatistiken, 469
Lambda
in Kreuztabellen, 292
Leerzeichen-getrennte Daten, 33
Leistung, 69
Zwischenspeichern von Daten, 69
Letzte
in "Mittelwerte", 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Levene-Test
in der Explorativen Datenanalyse, 287
in einfaktorieller ANOVA, 326
in GLM - Univariat, 338
Likelihood-Quotienten-Chi-Quadrat
in Kreuztabellen, 292
in Ordinale Regression, 368
Lilliefors-Test
in der Explorativen Datenanalyse, 287
Lineare Regression, 354
Auswahlmethoden für Variablen, 355, 363
Auswahlvariable, 357
Blöcke, 354
Diagramme, 357
Exportieren von Modellinformationen, 359
Fehlende Werte, 363
Gewichtungen, 354
Residuen, 359
Speichern von neuen Variablen, 359
Statistiken, 362
zusätzliche Funktionen beim Befehl, 364
Lineares Modell
in Kurvenanpassung, 374
Linearitätstests
in "Mittelwerte", 304
Liste der Zielvariablen, 489
Logarithmisches Modell
in Kurvenanpassung, 374
Logistisches Modell
in Kurvenanpassung, 374
Lokale Kodierung, 273
Löschen mehrerer EXECUTE-Befehle in Syntaxdateien, 274
Löschen von Ausgaben, 226
Lotus 1-2-3-Dateien, 15, 47, 508
Hinzufügen eines Menüeintrags zum Senden von Daten an Lotus, 508
öffnen, 15
speichern, 47
LSD nach Fisher
in GLM, 335

M-Schätzer
in der Explorativen Datenanalyse, 286
M-Schätzer nach Hampel
in der Explorativen Datenanalyse, 286
M-Schätzer nach Huber
in der Explorativen Datenanalyse, 286
Mahalanobis-Abstand
in der Diskriminanzanalyse, 385
in Lineare Regression, 359
Mann-Whitney-*U*-Test
in Tests bei zwei unabhängigen Stichproben, 427
Mantel-Haenszel-Statistik
in Kreuztabellen, 292
Maßeinheit, 491
Maximum
in der Explorativen Datenanalyse, 286
in Deskriptive Statistiken, 281
in Häufigkeiten, 276
in "Mittelwerte", 304
in OLAP-Würfel, 309
in Verhältnisstatistiken, 469
in Zusammenfassen, 299
Vergleichen von Berichtsspalten, 453
Maximum Likelihood
in der Faktorenanalyse, 392
McFadden-R^2
in Ordinale Regression, 368
McNemar-Test
in Kreuztabellen, 292
in Tests bei zwei verbundenen Stichproben, 429
Median
in der Explorativen Datenanalyse, 286
in Häufigkeiten, 276
in "Mittelwerte", 304
in OLAP-Würfel, 309
in Verhältnisstatistiken, 469
in Zusammenfassen, 299
Mediantest
in Tests bei zwei unabhängigen Stichproben, 431

mehrere Ansichten/Fenster
  Daten-Editor, 103
Mehrere geöffnete Datendateien, 105, 491
  Unterdrücken, 109
Mehrfachantworten
  zusätzliche Funktionen beim Befehl, 443
Mehrfachantworten: Häufigkeiten, 438
  Fehlende Werte, 438
Mehrfachantworten: Kreuztabellen, 440
  Definieren von Wertebereichen, 442
  Fehlende Werte, 442
  Paaren von Variablen aus verschiedenen Antworten-Sets, 442
  Prozentwerte basierend auf Antworten, 442
  Prozentwerte basierend auf Fällen, 442
  Prozentwerte für Zellen, 442
Mehrfachantworten-Sets
  definieren, 117
  mehrere Kategorien, 117
  Set aus dichotomen Variablen, 117
Mehrfachantworten-Sets definieren, 437
  Dichotomien, 437
  Kategorien, 437
  Set-Labels, 437
  Set-Namen, 437
Mehrfache Regression
  in Lineare Regression, 354
Mehrfachvergleiche
  in einfaktorieller ANOVA, 323
Menüs, 508
  Anpassen, 508
Messniveau, 84, 114
  definieren, 84
  Symbole in Dialogfeldern, 9
Metadateien, 232
  Exportieren von Diagrammen, 232
Metrisch, 84
  Messniveau, 84
Metrische Variablen
  Klassieren, um kategoriale Variablen zu erstellen, 131
Microsoft Access, 20
Minimum
  in der Explorativen Datenanalyse, 286
  in Deskriptive Statistiken, 281
  in Häufigkeiten, 276
  in "Mittelwerte", 304
  in OLAP-Würfel, 309
  in Verhältnisstatistiken, 469
  in Zusammenfassen, 299
  Vergleichen von Berichtsspalten, 453
Mittelwert
  in Bericht in Spalten, 452
  in Bericht in Zeilen, 447
  in der Explorativen Datenanalyse, 286
  in Deskriptive Statistiken, 281
  in einfaktorieller ANOVA, 326
  in Häufigkeiten, 276
  in "Mittelwerte", 304
  in OLAP-Würfel, 309
  in Verhältnisstatistiken, 469
  in Zusammenfassen, 299
  Untergruppe, 302, 307
  von mehreren Berichtsspalten, 453
Mittelwerte, 302
  Optionen, 304
  Statistiken, 304
Mittelwerte von Untergruppen, 302, 307
Modalwert
  in Häufigkeiten, 276
Modell kategorisieren
  in Ordinale Regression, 369
Modell skalieren
  in Ordinale Regression, 370
Modelldatei
  Laden von gespeicherten Modellen zum Bewerten von Daten, 181
mrInterview, 42
Multidimensionale Skalierung, 461
  Anzeigeoptionen, 465
  Beispiel, 461
  Definieren der Datenform, 463
  Dimensionen, 464
  Distanzmaße, 463
  Erstellen von Distanzmatrizen, 463
  Konditionalität, 464
  Kriterien, 465
  Messniveaus, 464
  Skalierungsmodelle, 464
  Statistiken, 461
  Transformieren von Werten, 463
  zusätzliche Funktionen beim Befehl, 466
Multipler Spannweitentest nach Ryan-Einot-Gabriel-Welsch
  in einfaktorieller ANOVA, 323
  in GLM, 335
Multiples *F* nach Ryan-Einot-Gabriel-Welsch
  in einfaktorieller ANOVA, 323
  in GLM, 335
Multiples *R*
  in Lineare Regression, 362
Multiplikation
  Multiplizieren über Berichtsspalten, 453
Musterdifferenzmaß
  in Distanzen, 351
Mustermatrix
  in der Faktorenanalyse, 389

Nagelkerke-R^2
  in Ordinale Regression, 368
Newman-Keuls
  in GLM, 335
Nicht standardisierte Residuen
  in GLM, 337

Nichtparametrische Tests
Chi-Quadrat, 417
Kolmogorov-Smirnov-Test bei einer Stichprobe, 424
Sequenzentest, 422
Tests bei mehreren unabhängigen Stichproben, 431
Tests bei mehreren verbundenen Stichproben, 434
Tests bei zwei unabhängigen Stichproben, 426
Tests bei zwei verbundenen Stichproben, 429
Nominal, 84
Messniveau, 84, 114
Normalrangwerte
beim Bilden der Rangfolge, 153
Normalverteilungsdiagramme
in der Explorativen Datenanalyse, 287
in Lineare Regression, 357
Numerisches Format, 84, 86

Öffnen von Dateien, 15–19, 33
Datendateien, 15–16
dBASE-Dateien, 15, 18
Excel-Dateien, 15, 17
Festlegen der standardmäßigen Datei-Speicherstellen, 504
Lotus 1-2-3-Dateien, 15
Stata-Dateien, 18
SYSTAT-Dateien, 15
Tabellenkalkulationsdateien, 15, 17
Tabulator-getrennte Dateien, 15
Textdatendateien, 33
OK (Schaltfläche), 8
OLAP-Würfel, 307
Statistiken, 309
Titel, 312
OMS, 523, 547
Ausgabe aus dem Viewer ausschließen, 534
Ausgabeobjekttypen, 526
Befehls-IDs, 527
SAV-Dateiformat, 529, 535
SPSS-Datendateiformat, 529, 535
Steuern von Pivots in Tabellen, 529, 540
Tabellenuntertypen, 527
Textformat, 529
Variablennamen in SAV-Dateien, 542
Verwenden von XSLT mit OXML, 548
XML, 529, 543
Online-Hilfe, 12
Statistik-Assistent, 11
Optionen, 491, 493–494, 496–497, 499, 502, 504–505
allgemein, 491
Beschriftung der Ausgabe, 497
Daten, 494
Diagramme, 499
Skripte, 505
Variablenansicht, 496
Verzeichnis für temporäre Dateien, 504
Viewer, 493
Vorlage für Pivot-Tabellen, 502
Währung, 496
zweistellige Jahresangaben, 494
Ordinal, 84
Messniveau, 84, 114
Ordinale Regression , 365
Modell kategorisieren, 369
Modell skalieren, 370
Optionen, 366
Statistiken, 365
Verknüpfung, 366
zusätzliche Funktionen beim Befehl, 371
OXML, 548

Paarweiser Vergleichstest nach Gabriel
in einfaktorieller ANOVA, 323
in GLM, 335
Paarweiser Vergleichstest nach Games und Howell
in einfaktorieller ANOVA, 323
in GLM, 335
Paralleles Modell
in der Reliabilitätsanalyse, 456, 458
Parallelitätstest für Linien
in Ordinale Regression, 368
Parameterschätzer
in GLM - Univariat, 338
in Ordinale Regression, 368
Partielle Diagramme
in Lineare Regression, 357
Partielle Korrelationen, 345
Fehlende Werte, 347
in Lineare Regression, 362
Korrelationen nullter Ordnung, 347
Optionen, 347
Statistiken, 347
zusätzliche Funktionen beim Befehl, 347
PDF
Exportieren von Ausgaben, 232, 235
Pearson-Chi-Quadrat
in Kreuztabellen, 292
in Ordinale Regression, 368
Pearson-Korrelation
in bivariaten Korrelationen, 341
in Kreuztabellen, 292
Pearson-Residuen
in Ordinale Regression, 368
Perzentile
in der Explorativen Datenanalyse, 286
in Häufigkeiten, 276
Phi
in Kreuztabellen, 292
Phi-Quadrat-Distanzmaß
in Distanzen, 351
Pivot-Tabellen, 225, 231–232, 240, 245–248, 251–252, 254–256, 258–259, 265–268, 502
allgemeine Eigenschaften, 254
Ändern der Anzeigereihenfolge, 246
Ändern der Vorlage, 252

Anpassen an die Seitengröße, 254, 259
Anpassen der Standard-Spaltenbreite, 502
Anzeigen ausgeblendeter Rahmen, 265
Aufheben der Gruppierung von Zeilen oder Spalten, 247
Ausblenden, 225
Ausrichtung, 262
Auswählen von Zeilen und Spalten, 266
bearbeiten, 245
Drehen von Beschriftungen, 248
Drucken umfangreicher Tabellen, 267
Drucken von Schichten, 240
Eigenschaften, 254
Ein- und Ausblenden von Zellen, 251
Einfügen als Tabellen, 231
Einfügen in andere Anwendungen, 231
Einfügen von Gruppenbeschriftungen, 247
Erklärungen, 263–264
Erstellen aus Pivot-Tabellen, 268
exportieren als HTML, 232
Festlegen von Tabellenumbrüchen, 267
Fortsetzungstext, 259
Fußnoten, 263–264
Fußnoteneigenschaften, 255
Gitterlinien, 265
Gruppieren von Zeilen oder Spalten, 247
Hintergrundfarbe, 261
kopieren in andere Anwendungen, 231
Löschen von Gruppenbeschriftungen, 247
pivotieren, 245
Rahmen, 258
Ränder, 262
Schichten, 248
Schriftarten, 261
Standardvorlage für neue Tabellen, 502
Transponieren von Zeilen und Spalten, 247
Verschieben von Zeilen und Spalten, 246
Verwenden von Symbolen, 245
Zellenbreiten, 265
Zelleneigenschaften, 261–262
Zellenformate, 256
pivotieren
mit OMS für exportierte Ausgabe steuern, 540
PLUM
in Ordinale Regression, 365
PNG-Dateien, 232, 239
Exportieren von Diagrammen, 232, 239
Polynomiale Kontraste
in einfaktorieller ANOVA, 322
in GLM, 332–333
portable Dateien
Variablennamen, 47
Portnummern, 73
Post-Hoc-Mehrfachvergleiche, 323
PostScript-Dateien (Encapsulated), 232, 239
Exportieren von Diagrammen, 232, 239
Potenzmodell
in Kurvenanpassung, 374
PowerPoint, 235
Exportieren von Ausgaben als PowerPoint, 235
PowerPoint-Format
Exportieren von Ausgaben, 232
Preisbezogenes Differential (PRD)
in Verhältnisstatistiken, 469
Produktionsjobs, 514, 519–521
Ausführen mehrerer Produktionsjobs, 521
Ausgabedateien, 514
Befehlszeilenschalter, 521
Ersetzen von Werten in Syntaxdateien, 519
Exportieren von Diagrammen, 514
Konvertieren von Produktionsmodus-Dateien, 522
Planen von Produktionsjobs, 521
Syntaxregeln, 514
Profilplots
in GLM, 334
Prognose
in Kurvenanpassung, 375
Programmieren mit Befehlssprache, 269
Prozentwerte
in Kreuztabellen, 295
Punktformat (DOT), 84, 86

Q nach R-E-G-W
in einfaktorieller ANOVA, 323
in GLM, 335
Quadratisches Modell
in Kurvenanpassung, 374
Quadratsumme, 331
in GLM, 330
Quadrierte Euklidische Distanz
in Distanzen, 351
Quancept, 42
Quanvert, 42
Quartile
in Häufigkeiten, 276
Quartimax-Rotation
in der Faktorenanalyse, 394

r-Korrelationskoeffizient
in bivariaten Korrelationen, 341
in Kreuztabellen, 292
R-Statistik
in Lineare Regression, 362
in “Mittelwerte”, 304
R^2
Änderung in R^2, 362
in Lineare Regression, 362
in “Mittelwerte”, 304
Rahmen, 258, 265
Anzeigen ausgeblendeter Rahmen, 265
Rand-Homogenitätstest
in Tests bei zwei verbundenen Stichproben, 429
Rang-Korrelationskoeffizient
in bivariaten Korrelationen, 341
Rankit-Schätzungen, 153

Rao-*V*
in der Diskriminanzanalyse, 385
Rauschverarbeitung
in der Two-Step-Clusteranalyse, 402
Rechtschreibung, 96
Datenlexikon, 494
Reduzieren von Kategorien, 131
Referenzkategorie
in GLM, 332–333
Regression
Diagramme, 357
Lineare Regression, 354
Mehrfache Regression, 354
Regression mit partiellen kleinsten Quadraten, 376
Exportieren von Variablen, 379
Modell, 378
Regressionskoeffizienten
in Lineare Regression, 362
Relatives Risiko
in Kreuztabellen, 292
Reliabilitätsanalyse, 456
ANOVA-Tabelle, 458
Beispiel, 456
deskriptive Statistiken, 458
Hotellings T^2, 458
Inter-Item-Korrelationen und -Kovarianzen, 458
Korrelationskoeffizienten in Klassen, 458
Kuder-Richardson-20, 458
Statistiken, 456, 458
Tukeys Additivitätstest, 458
zusätzliche Funktionen beim Befehl, 460
Remote-Server, 72–73, 76, 78
Anmelden beim, 72
bearbeiten, 73
hinzufügen, 73
Relative Pfade, 78
verfügbare Prozeduren, 78
Zugriff auf Datendateien, 76
Residuen
in Kreuztabellen, 295
Speichern in Kurvenanpassung, 375
Speichern in Lineare Regression, 359
Residuen-Diagramme
in GLM - Univariat, 338
Rho
in bivariaten Korrelationen, 341
in Kreuztabellen, 292
Risiko
in Kreuztabellen, 292
ROC-Kurve, 471
Statistiken und Diagramme, 472
Rückwärtselimination
in Lineare Regression, 355

S-Modell
in Kurvenanpassung, 374
S-Streß
in Multidimensionale Skalierung, 461
SAS-Dateien
öffnen, 15
speichern, 47
SAV-Dateiformat
Ausgabe an eine SPSS-Datendatei weiterleiten, 529, 535
Savage-Werte, 153
Schätzer der Effektgröße
in GLM - Univariat, 338
Schätzer der Schärfe
in GLM - Univariat, 338
Scheffé-Test
in einfaktorieller ANOVA, 323
in GLM, 335
Schichten, 240, 248, 250, 254, 259
anzeigen, 248, 250
Drucken, 240, 254, 259
erstellen, 248
in Kreuztabellen, 292
in Pivot-Tabellen, 248
Schiefe
in Bericht in Spalten, 452
in Bericht in Zeilen, 447
in der Explorativen Datenanalyse, 286
in Deskriptive Statistiken, 281
in Häufigkeiten, 276
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Schlüsseltabelle, 192
Schriftarten, 103, 228, 261
im Daten-Editor, 103
im Gliederungsfenster, 228
Schrittweise Auswahl
in Lineare Regression, 355
Seiteneinrichtung, 241, 243
Diagrammgröße, 243
Kopf- und Fußzeilen, 241
Seiteneinstellung
in Berichte in Spalten, 454
in Berichte in Zeilen, 448
Seitennumerierung, 243
in Berichte in Spalten, 454
in Berichte in Zeilen, 448
Seitenverhältnis, 499
Sequenzentest
Fehlende Werte, 423
Optionen, 423
Statistiken, 423
Trennwerte, 422–423
zusätzliche Funktionen beim Befehl, 424
Server, 72–73
Anmelden beim, 72
bearbeiten, 73
hinzufügen, 73
Namen, 73

Portnummern, 73
Shapiro-Wilk-Test
in der Explorativen Datenanalyse, 287
Sidak-*T*-Test
in einfaktorieller ANOVA, 323
in GLM, 335
Sitzungs-Journal, 504
Skala
in der Reliabilitätsanalyse, 456
in Multidimensionale Skalierung, 461
Messniveau, 114
Skalierung
Pivot-Tabellen, 254, 259
Skripte
Standardsprache, 505
Skripts, 508, 512
Ausführen mit Symbolleisten-Schaltflächen, 512
Hinzufügen zu Menüs, 508
Somers-*d*
in Kreuztabellen, 292
sortieren
Variablen, 186
Sortieren von Fällen, 185
Sortieren von Variablen, 186
Spalten, 265–266
Ändern der Breite in Pivot-Tabellen, 265
Auswählen in Pivot-Tabellen, 266
Spaltenbreite, 89, 254, 265, 502
Beeinflussen der Breite bei Texten mit Zeilenumbrüchen, 254
Einstellen der Standardbreite, 502
Festlegen der maximalen Breite, 254
im Daten-Editor, 89
Pivot-Tabellen, 265
Spaltenprozente
in Kreuztabellen, 295
Spearman-Brown-Reliabilität
in der Reliabilitätsanalyse, 458
Speichern der Ausgabe, 232, 235, 237
Excel-Format, 232, 234
HTML, 232, 234
HTML-Format, 232
PDF-Format, 232, 235
PowerPoint-Format, 232, 235
Textformat, 232, 237
Word-Format, 232, 234
Speichern von Dateien, 46–47
Datenbankdatei-Abfragen, 32
Datendateien, 47
Festlegen der standardmäßigen Datei-Speicherstellen, 504
SPSS-Datendateien, 46
Speichern von Diagrammen, 232, 238–239
BMP-Dateien, 232, 239
EMF-Dateien, 232
EPS-Dateien, 232, 239
JPEG-Dateien, 232, 238
Metadateien, 232
PICT-Dateien, 232
PNG-Dateien, 239
PostScript-Dateien, 239
TIFF-Dateien, 239
Speicherplatz, 68–69
temporär, 68–69
Speicherzuweisung
in der Two-Step-Clusteranalyse, 402
Spezielle Währungsformate, 84, 496
Split-Half-Reliabilität
in der Reliabilitätsanalyse, 456, 458
SPP-Dateien
Konvertieren in SPJ-Dateien, 522
Sprache
Ändern der Ausgabesprache, 491
SPSS-Datendateiformat
Ausgabe an eine Datendatei weiterleiten, 529, 535
SPSS-Produktionsmodus, 491
Konvertieren von Dateien in Produktionsjobs, 522
Verwenden von Befehlssyntax aus der Journaldatei, 491
SPSSTMPDIR (Umgebungsvariable), 504
Standardabweichung
in Bericht in Spalten, 452
in Bericht in Zeilen, 447
in der Explorativen Datenanalyse, 286
in Deskriptive Statistiken, 281
in GLM - Univariat, 338
in Häufigkeiten, 276
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Verhältnisstatistiken, 469
in Zusammenfassen, 299
Standardfehler
in der Explorativen Datenanalyse, 286
in Deskriptive Statistiken, 281
in GLM, 337–338
in Häufigkeiten, 276
in ROC-Kurve, 472
Standardfehler der Kurtosis
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Standardfehler der Schiefe
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Standardfehler des Mittelwerts
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Standardisierte Residuen
in GLM, 337
in Lineare Regression, 359
Standardisierte Werte
in Deskriptive Statistiken, 280

Standardisierung
in der Two-Step-Clusteranalyse, 402
Standardmäßige Datei-Speicherstellen, 504
Startwert für Zufallszahlen, 144
Stata-Dateien, 18
einlesen, 15
öffnen, 18
speichern, 47
Statistik-Assistent, 11
Statusleiste, 6
Stengel-Blatt-Diagramme
in der Explorativen Datenanalyse, 287
Stichprobe
Zufallsstichprobe, 203
Streng paralleles Modell
in der Reliabilitätsanalyse, 456, 458
Streß
in Multidimensionale Skalierung, 461
Streudiagramme
in Lineare Regression, 357
Streuungskoeffizient (COD)
in Verhältnisstatistiken, 469
Streuungsmaße
in der Explorativen Datenanalyse, 286
in Deskriptive Statistiken, 281
in Häufigkeiten, 276
in Verhältnisstatistiken, 469
String-Format, 84
String-Variablen, 88, 97
Berechnen von neuen String-Variablen, 142
Eingeben von Daten, 97
Fehlende Werte, 88
in Dialogfeldern, 7
lange String-Variablen in früheren Versionen aufteilen, 47
umkodieren in fortlaufende Ganzzahlen, 155
Student-Newman-Keuls-Prozedur
in einfaktorieller ANOVA, 323
in GLM, 335
Student-*T*-Test, 313
Studentisierte Residuen
in Lineare Regression, 359
Studie mit Fallkontrolle
T-Test bei gepaarten Stichproben, 316
Studie mit zugeordneten Paaren
in T-Test bei gepaarten Stichproben, 316
Suchen und Ersetzen
Viewer-Dokumente, 229
Summe
in Deskriptive Statistiken, 281
in Häufigkeiten, 276
in “Mittelwerte”, 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Symbole
in Dialogfeldern, 9
Symbolleisten, 509, 511–512
Anpassen, 509, 511
Anzeigen in verschiedenen Fenstern, 511
ein- und ausblenden, 509
erstellen, 509, 511
Erstellen von neuen Symbolen, 512
Syntax, 269, 273, 511, 514
ausführen, 273
Ausführen von Befehlssyntax mit Symbolleisten-Schaltflächen, 511
Ausgabe-Log, 271
einfügen, 271
Journaldatei, 274
Produktionsjobs, Regeln, 514
Syntaxregeln, 269
Unicode-Befehlssyntaxdateien, 273
Zugreifen auf *SPSS Command Syntax Reference*, 12
SYSTAT-Dateien, 15
öffnen, 15

T-Test
in GLM - Univariat, 338
in T-Test bei einer Stichprobe, 318
in T-Test bei gepaarten Stichproben, 316
in T-Test bei unabhängigen Stichproben, 313
T-Test bei einer Stichprobe, 318
Fehlende Werte, 319
Konfidenzintervalle, 319
Optionen, 319
zusätzliche Funktionen beim Befehl, 320
T-Test bei gepaarten Stichproben, 316
Auswählen von gepaarten Variablen, 316
Fehlende Werte, 317
Optionen, 317
T-Test bei unabhängigen Stichproben, 313
Fehlende Werte, 316
Gruppen definieren, 315
Gruppenvariablen, 315
Konfidenzintervalle, 316
Optionen, 316
String-Variablen, 315
T-Test bei zwei Stichproben
in T-Test bei unabhängigen Stichproben, 313
T2 nach Tamhane
in einfaktorieller ANOVA, 323
in GLM, 335
T3 nach Dunnett
in einfaktorieller ANOVA, 323
in GLM, 335
T4253H-Glättung, 176
Tabellen, 267
Ausrichtung, 262
Festlegen von Tabellenumbrüchen, 267
Hintergrundfarbe, 261
Ränder, 262
Schriftarten, 261
Zelleneigenschaften, 261–262

Tabellendiagramm, 268
Tabellenkalkulationsdateien, 15, 17, 49
  Einlesen von Bereichen, 17
  Einlesen von Variablennamen, 17
  öffnen, 17
  Schreiben von Variablennamen, 49
Tabellenumbrüche, 267
Tabellenuntertypen, 527
  im Vergleich mit Beschriftungen, 528
Tabellenvorlagen, 252–253
  erstellen, 253
  zuweisen, 252
Tabulator-getrennte Dateien, 15, 17, 33, 47, 49
  Einlesen von Variablennamen, 17
  öffnen, 15
  Schreiben von Variablennamen, 49
  speichern, 47
Tau-*b*
  in Kreuztabellen, 292
Tau-*c*
  in Kreuztabellen, 292
Teilmengen von Fällen
  auswählen, 200, 202, 204
  Zufallsstichprobe, 203
Temporäre aktive Datei, 69
Temporärer Speicherplatz, 68–69
Test auf Binomialverteilung, 420
  Dichotomien, 420
  Fehlende Werte, 421
  Optionen, 421
  Statistiken, 421
  zusätzliche Funktionen beim Befehl, 422
Test Extremreaktionen nach Moses
  in Tests bei zwei unabhängigen Stichproben, 427
Tests auf Homogenität der Varianzen.
  in einfaktorieller ANOVA, 326
  in GLM - Univariat, 338
Tests auf Normalverteilung
  in der Explorativen Datenanalyse, 287
Tests auf Unabhängigkeit
  Chi-Quadrat, 292
Tests bei mehreren unabhängigen Stichproben, 431
  Definieren des Bereichs, 433
  Fehlende Werte, 433
  Gruppenvariablen, 433
  Optionen, 433
  Statistiken, 433
  Testtypen, 432
  zusätzliche Funktionen beim Befehl, 434
Tests bei mehreren verbundenen Stichproben, 434
  Statistiken, 435
  Testtypen, 435
  zusätzliche Funktionen beim Befehl, 435
Tests bei zwei unabhängigen Stichproben, 426
  Fehlende Werte, 428
  Gruppen definieren, 428
  Gruppenvariablen, 428
  Optionen, 428
  Statistiken, 428
  Testtypen, 427
  zusätzliche Funktionen beim Befehl, 429
Tests bei zwei verbundenen Stichproben, 429
  Fehlende Werte, 431
  Optionen, 431
  Statistiken, 431
  Testtypen, 430
  zusätzliche Funktionen beim Befehl, 431
Text, 33, 228–229, 232, 237
  Datendateien, 33
  einfügen im Viewer, 228
  Exportieren von Ausgaben als Text, 232, 237
  Hinzufügen von Textdateien im Viewer, 229
TIFF-Dateien, 239
  Exportieren von Diagrammen, 232, 239
Titel, 228
  Diagramme, 481
  einfügen im Viewer, 228
  in OLAP-Würfel, 312
Toleranz
  in Lineare Regression, 362
Transformationsmatrix
  in der Faktorenanalyse, 389
Transponieren von Variablen und Fällen, 188
Transponieren von Zeilen und Spalten, 247
Trendbereinigte Normalverteilungsdiagramme
  in der Explorativen Datenanalyse, 287
Tukey-*B*-Test
  in einfaktorieller ANOVA, 323
  in GLM, 335
Tukey-Biweight-Schätzer
  in der Explorativen Datenanalyse, 286
Tukey-Schätzungen, 153
Tukeys Additivitätstest
  in der Reliabilitätsanalyse, 456, 458
Two-Step-Clusteranalyse, 399
  Diagramme, 404
  in Arbeitsdatei speichern, 405
  in externer Datei speichern, 405
  Optionen, 402
  Statistiken, 405

Umbenennen von Daten-Sets, 108
Umgebungsvariablen, 504
  SPSSTMPDIR, 504
Umkodieren von Werten, 131, 147–148, 150–151, 155
Umordnen von Zeilen und Spalten, 246
Umstrukturieren von Daten, 206, 209–211, 213–216, 218–222
  Auswählen von Daten für “Fälle zu Variablen”, 219
  Auswählen von Daten für “Variablen zu Fälle”, 211
  Beispiel für einen Index bei “Variablen zu Fälle”, 214
  Beispiel für “Fälle zu Variablen”, 210
  Beispiel für “Variablen zu Fälle”, 209
  Beispiel für zwei Indizes bei “Variablen zu Fälle”, 215

Erstellen einer einzelnen Indexvariable für "Variablen zu Fälle", 215
Erstellen mehrerer Indexvariablen für "Variablen zu Fälle, 216
Erstellen von Indexvariablen für "Variablen zu Fälle", 213
Optionen für "Variablen zu Fälle", 218, 221
Sortieren von Daten für "Fälle zu Variablen", 220
Übersicht, 206
Umstrukturierungstypen, 206
und gewichtete Daten, 222
Variablengruppen für "Variablen zu Fälle", 210
Unähnlichkeitsmaße nach Lance und Williams, 351
in Distanzen, 351
Ungewichtete kleinste Quadrate
in der Faktorenanalyse, 392
Unicode, 15, 46
Unicode-Befehlssyntaxdateien, 273
Unsicherheitskoeffizient
in Kreuztabellen, 292
Untertitel
Diagramme, 481
Untertypen, 527
im Vergleich mit Beschriftungen, 528

V
in Kreuztabellen, 292
Van der Waerden-Schätzungen, 153
Variablen, 9, 82, 99–100, 206, 485–486, 491
Anzeigereihenfolge in Dialogfeldern, 491
definieren, 82
Definieren von Variablen-Sets, 486
Einfügen von neuen Variablen, 99
in Dialogfeldern, 7
in Dialogfeldern auswählen, 9
Informationen zu Variablen in Dialogfeldern, 10
Informationen zur Definition, 485
sortieren, 186
Suchen im Daten-Editor, 100
Umbenennen für zusammengefügte Datendateien, 192
umkodieren, 147–148, 150–151, 155
Umstrukturieren in Fälle, 206
Verschieben, 100
Variablen-Sets, 486–487
definieren, 486
verwenden, 487
Variablenansicht, 81
Anpassen, 95, 496
Variablenattribute, 89–90
Benutzerdefiniert, 91
Kopieren und Einfügen, 89–90
Variablenbeschreibung, 485
Variablenlabels, 87, 491, 497
Einfügen von Zeilenumbrüchen, 88
im Gliederungsfenster, 497
in Dialogfeldern, 7, 491
in Pivot-Tabellen, 497
in zusammengefügten Datendateien, 192
Variablenlisten, 489
Umsortieren von Listen der Zielvariablen, 489
Variablennamen, 83, 491
durch das OMS erzeugt, 542
in Dialogfeldern, 7, 491
lange Variablennamen in früheren Versionen kürzen, 47
portable Dateien, 47
Regeln, 83
Variablennamen für gemischte Fälle, 83
Zeilenumbruch bei langen Variablennamen in der Ausgabe, 83
Variablenpaare, 206
erstellen, 206
Varianz
in Bericht in Spalten, 452
in Bericht in Zeilen, 447
in der Explorativen Datenanalyse, 286
in Deskriptive Statistiken, 281
in Häufigkeiten, 276
in "Mittelwerte", 304
in OLAP-Würfel, 309
in Zusammenfassen, 299
Varianz-Inflationsfaktor
in Lineare Regression, 362
Varianzanalyse
in einfaktorieller ANOVA, 321
in Kurvenanpassung, 372
in Lineare Regression, 362
in "Mittelwerte", 304
Variationskoeffizient (COV)
in Verhältnisstatistiken, 469
Varimax-Rotation
in der Faktorenanalyse, 394
Verallgemeinerte kleinste Quadrate
in der Faktorenanalyse, 392
Verarbeitung von aufgeteilten Dateien, 199
Verbergen (Ausschließen) der Ausgabe aus dem Viewer bei OMS, 534
Verbundene Stichproben, 429, 434
Vergleichen von Gruppen
in OLAP-Würfel, 311
Vergleichen von Variablen
in OLAP-Würfel, 311
Verhältnisstatistik, 467
Statistiken, 469
Verknüpfung
in Ordinale Regression, 366
Verschieben von Zeilen und Spalten, 246
Verteilter Modus, 72–73, 76, 78
Relative Pfade, 78
verfügbare Prozeduren, 78
Zugriff auf Datendateien, 76
Verteilungsmaße
in Deskriptive Statistiken, 281
in Häufigkeiten, 276
Vertikaler Beschriftungstext, 248

Verzeichnis für temporäre Dateien, 504
  Festlegen des Speicherorts bei lokalen Analysen, 504
  SPSSTMPDIR (Umgebungsvariable), 504
Viewer, 224–228, 243–244, 493, 497
  Abstand zwischen Ausgabeobjekten, 243
  Ändern der Größe von Gliederungsobjekten, 228
  Ändern der Schriftart in der Gliederung, 228
  Ändern von Gliederungsebenen, 227
  Anzeigen von Datenwerten, 497
  Anzeigen von Variablenlabels, 497
  Anzeigen von Variablennamen, 497
  Anzeigen von Wertelabels, 497
  Anzeigeoptionen, 493
  ausblenden von Ergebnissen, 225
  Ausgabetypen bei OMS ausschließen, 534
  Ergebnisfenster, 224
  Erweitern der Gliederung, 227
  Gliederung, 226
  Gliederungsfenster, 224
  Löschen von Ausgaben, 226
  Reduzieren der Gliederung, 227
  Speichern von Dokumenten, 244
  Suchen und Ersetzen von Informationen, 229
  Verschieben der Ausgabe, 225
virtuelle aktive Datei, 68
Vorhergesagte Werte
  Speichern in Kurvenanpassung, 375
  Speichern in Lineare Regression, 359
Vorhersageintervalle
  Speichern in Kurvenanpassung, 375
  Speichern in Lineare Regression, 359
Vorlagen, 89–90, 482, 499
  Diagramme, 482
  in Diagrammen, 499
  Variablendefinition, 89–90
  Verwenden einer externen Datendatei als Vorlage, 120
Vorlauffunktion, 176
Vorwärtsselektion
  in Lineare Regression, 355
Vorzeichentest
  in Tests bei zwei verbundenen Stichproben, 429

Wachstumsmodell
  in Kurvenanpassung, 374
Währungsformate, 496
Wald-Wolfowitz-Sequenzen
  in Tests bei zwei unabhängigen Stichproben, 427
Waller-Duncan-*T*-Test
  in einfaktorieller ANOVA, 323
  in GLM, 335
Wechselwirkungsterme, 331, 371
Welch-Statistik
  in einfaktorieller ANOVA, 326
Wertelabels, 87, 97, 103, 111, 497
  Einfügen von Zeilenumbrüchen, 88
  im Daten-Editor, 103
  im Gliederungsfenster, 497
  in Pivot-Tabellen, 497
  in zusammengefügten Datendateien, 192
  Kopieren, 116
  mehreren Variablen zuweisen, 116
  Speichern in Excel-Dateien, 47
  Verwenden bei der Dateneingabe, 97
Wichtigkeitsdiagramm
  in der Two-Step-Clusteranalyse, 404
Wichtigkeitsdiagramme für Variablen
  in der Two-Step-Clusteranalyse, 404
Wiederholte Kontraste
  in GLM, 332–333
Wilcoxon-Test
  in Tests bei zwei verbundenen Stichproben, 429
Wilks-Lambda
  in der Diskriminanzanalyse, 385
Wissenschaftliche Notation, 84, 491
  unterdrücken in der Ausgabe, 491
Word-Format
  Breite Tabellen, 232
  Exportieren von Ausgaben, 232, 234

XML
  Ausgabe an XML weiterleiten, 529
  OXML-Ausgabe aus OMS, 548
  Speichern von Ausgaben als XML, 523
  Tabellenstruktur in OXML, 543
XSLT
  Verwenden mit OXML, 548

Z-Werte
  beim Bilden der Rangfolge, 153
  in Deskriptive Statistiken, 280
  Speichern als Variablen, 280
Zählen von Werten, 145
Zeilen, 266
  Auswählen in Pivot-Tabellen, 266
Zeilenprozente
  in Kreuztabellen, 295
Zeilenumbrüche, 254
  Beeinflussen der Spaltenbreite bei Texten mit Zeilenumbrüchen, 254
  Variablen- und Wertelabels, 88
Zeitreihenanalyse
  Prognose, 375
  Vorhersagen von Fällen, 375
Zeitreihendaten
  Datentransformationen, 173
  Definieren von Datumsvariablen, 174
  Ersetzen fehlender Werte, 178
  Erstellen neuer Zeitreihenvariablen, 175
  Transformationsfunktionen, 176
Zellen in Pivot-Tabellen, 256, 264–265
  Ausblenden, 251
  Breiten, 265
  Einblenden, 251

Formate, 256
Zelleneigenschaften, 261–262
Zentrieren der Ausgabe, 226, 493
Zufallsstichprobe, 25
auswählen, 203
Datenbanken, 25
Startwert für Zufallszahlen, 144
Zurücksetzen (Schaltfläche), 8
Zusammenfassen, 297
Optionen, 299
Statistiken, 299
Zusammenfügen von Datendateien
Dateien mit unterschiedlichen Fällen, 189
Dateien mit unterschiedlichen Variablen, 192
Informationen aus dem Datenlexikon, 192
Umbenennen von Variablen, 192
Zusammengesetztes Wachstumsmodell
in Kurvenanpassung, 374
Zusammenhang linear-mit-linear
in Kreuztabellen, 292
Zwischenergebnisse
in Berichte in Spalten, 454